10-15-2018 9:00 AM

A Shared Frailty Competing Risk Model With Time Varying Covariates For Estimation Of Breast And Ovarian Cancer Risks In BRCA1/2 Families

Hae Young Jung
The University of Western Ontario

Graduate Program in Epidemiology and Biostatistics
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

Abstract

Mammographic screening (MS) and prophylactic surgery (PS) can potentially reduce cancer risks in *BRCA1/2* mutation families. The evaluation of these interventions is complicated both by the presence of competing events and by their values changing over time. We propose a competing risks model and provide cause-specific penetrance estimates to account for time-varying covariates in the presence of competing events. A shared frailty model is specified to account for familial residual dependence and is estimated to correct for the ascertainment bias induced by the sampling of families through probands via an ascertainment corrected likelihood approach. We evaluate the performance of the estimators for the parameters in the model and the plug-in estimators for the penetrance functions via simulations. We apply our proposed method to *BRCA1/2* mutation carrier families recruited through the Breast Cancer Family Registries and evaluate the effects of PS and MS on breast and ovarian cancer risks.

Keywords: competing risk model; time varying covariate; frailty; cancer screening; recurrent data; time-to-event data; family data.

Acknowlegements

First, I sincerely appreciate my supervisor Dr. Yun-hee Choi of the Department of Epidemiology and Biostatistics at Western Ontario University and Dr. Laurent Briollais of Lunenfeld-Tanenbaum Research Institute at Mount Sinai hospital for guiding me and being patient with my progress throughout my research. I was barely able to report desciptive statistics on the dataset in the beginning of my study but they taught me so much in every aspect of independent research. My growth is solely attributable to them. My only regret is that I could not add more than that I did because of my limitations in skillset and understanding.

For his valuable feeback on the manuscript, I am grateful to Dr. Guangyong Zou of the Department of Epidemiology and Biostatistics at Western Ontario University.

I thank Mount Sinai Hospital for permission to use the Breast Cancer Family Registry data for my analysis. I also want to acknowledge both postdoctoral Fellow Fode Tounkara for helping me throughout this research and Apostolos Dimitromanolakis for teaching me valuable skills in computation.

Lastly, I thank God and my family for believing in and supporting me whichever path I take. It could not have been possible to complete my thesis without their unconditional love.

Contents

Abstract i

Acknowlegements ii

List of Figures v

List of Tables ix

List of Appendices xi

List of Abbreviations xi

1 Introduction 1

1.1 Objectives 3

1.2 Organization of the thesis 4

2 Literature Review 5

2.1 Competing risk models 5

2.2 Frailty models for clustered time-to-event data 9

2.2.1 Shared frailty model in non-competing risks setting 10

2.2.2 Shared frailty model in competing risk setting 11

2.3 Time varying covariates 13

2.4 Ascertainment bias and ascertainment corrected likelihood 15

3 Statistical Models 17

3.1 Shared frailty competing risk model with time-varying covariates 17

3.2 Likelihood construction with ascertainment correction 20

3.3 Cause-specific penetrance function with time-varying covariates 23
3.4 Variance estimation 25

4 Simulation **27**
4.1 Objectives of the simulation study 27
4.2 Data generation 28
4.3 Simulation settings and evaluation criteria 31
4.4 Simulation results and discussions 34
4.4.1 Cause-specific penetrances 34
4.4.2 Model parameters 35

5 Application to hereditary breast and ovarian cancer families **46**
5.1 The data 46
5.2 Model specification 54
5.2.1 Effects of bilateral oophorectomy and mammographic screenings 54
5.2.2 Missing mutation status 56
5.3 Analysis results 57
5.3.1 Relative risk 57
5.3.2 *BRCA1* penetrance estimations 59
5.3.3 *BRCA2* penetrance estimations 63
5.3.4 Summary 67

6 Discussion **68**
6.1 Summary 68
6.2 Limitation and further work 69

A Additional simulation results **72**

B R codes **84**

Bibliography **91**

List of Figures

2.1 Effect of time-varying covariate over time in Permanent Exposure (PE), Exponential Decay (ED) and Cox and Oakes (CO) models given $t_x = 35$, $\beta = 1.5$, $\eta_0 = 0.5$ and $\eta_1 = 0.3$. 15

4.1 Accuracy and precision of the penetrance estimates for mutation carriers from n = 1000 families based on 500 simulations. The red line represents the true penetrance of the screened group and the blue line represents non-screened group. 40

4.2 Accuracy and precision of the penetrance estimates for mutation carriers from n = 500 families based on 500 simulations. The red line represents the true penetrance of the screened group and the blue line represents non-screened group. 41

4.3 Accuracy and precision of the penetrance estimates for mutation carriers from n = 250 families based on 500 simulations. The red line represents the true penetrance of the screened group and the blue line represents non-screened group. 42

4.4 Bias and precision of the parameter estimates for screen effect, β_s, expressed as bias $\pm$ 1.96ASE, based on 500 simulations. True value of the β_s is 0.668 for PE, 1.872 for ED and 3.401 for CO. 43

4.5 Bias and precision of the parameter estimates for mutation effect for event 1, $\beta_{g,1}$, expressed as bias $\pm$ 1.96ASE, based on 500 simulations. True value of the $\beta_{g,1}$ is 1.952 for PE, 1.858 for ED and 2.078 for CO. 44

4.6 Bias and precision of the parameter estimates for frailty parameter for event 1, k_1, expressed as bias $\pm$ 1.96ASE, based on 500 simulations. True value of the $\log(k_1)$ is 2.30 for low, 1.25 for medium and 0 for high familial dependence. . . 45

5.1 Total number of women entered the study with competing risks, *BRCA1/2* combined. 47
5.2 (*BRCA1*) Penetrance estimations for breast cancer (left panel) and ovarian cancer (right panel) conditioning no screening activities and surgeries. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals. 60
5.3 (*BRCA1*) Breast cancer (BC) penetrance estimations with the first mammographic screenings. The left most plot represents BC penetrance with no screening. To the right, they describe penetrance estimation with the first screening at age 35, 40 and 45, respectively. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals. 61
5.4 (*BRCA1*) Breast cancer (BC) penetrance estimations with three mammographic screenings. The left most plot represents BC penetrance with no screening. To the right, they describe penetrance estimations with the first screening at age 30 and 35 with the consecutive screening gap times of 5 years. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals. 62
5.5 (*BRCA1*) Breast cancer (BC) penetrance estimations with bilateral oophorectomy. The left most plot represents BC penetrance with no BO. To the right, they describe penetrance estimation with the BO at age 40, 45 and 50. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals. 63
5.6 (*BRCA2*) Penetrance estimations for breast cancer (left panel) and ovarian cancer (right panel) conditioning no screening activities and surgeries. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals. 64

5.7 (*BRCA2*) Breast cancer (BC) penetrance estimations with the first mammographic screenings. The left most plot represents BC penetrance with no screening. To the right, they describe penetrance estimation with the first screening at age 35, 40 and 45, respectively. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals. 65

5.8 (*BRCA2*) Breast cancer (BC) penetrance estimations with three mammographic screenings. The left most plot represents BC penetrance with no screening. To the right, they describe penetrance estimations with the first screening at age 30 and 35 with the consecutive screening gap times of 5 years. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals. 66

5.9 (*BRCA2*) Breast cancer (BC) penetrance estimations with bilateral oophorectomy. The left most plot represents BC penetrance with no BO. To the right, they describe penetrance estimation with the BO at age 40, 45 and 50. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals. 67

A.1 (1000 families) Boxplot of bias of the model parameters for PE TVC model under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1. 79

A.2 (1000 families) Boxplot of bias of the model parameters for ED TVC model under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1. 80

A.3 (1000 families) Boxplot of bias of the model parameters for CO TVC model under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1. 80

A.4 (1000 families) Boxplot of bias of the frailty parameters for PE TVC model under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1. 81

A.5 (1000 families) Boxplot of bias of the frailty parameters for ED TVC model under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1. 81
A.6 (1000 families) Boxplot of bias of the frailty parameters for CO TVC model under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1. 82
A.7 Bias and precision of the parameter estimates for mutation effect parameter for event 2, $\beta_{g,1}$, expressed as mean +/- 1.96ASE, based on 500 simulations. True value of the $\beta_{g,2}$ is 1.194 for PE, 1.224 for ED and 1.566 for CO. 83

List of Tables

2.1 Several possible choices of baseline hazard function 9

4.1 Parameter values used for generating family data for our simulation study and the corresponding penetrance values based on the true parameter values. 32

4.2 Summary of evaluation criteria. 33

4.3 (500 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Permanent Exposure TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations. 37

4.4 (500 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Exponential Decay TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations. 38

4.5 (500 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Cox and Oakes TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations. 39

5.1 Descriptive statistics of *BRCA1* positive families. 48

5.2 Descriptive statistics of *BRCA2* mutation positive analysis cohorts. 49

5.3 Descriptive statistics for the probands of *BRCA1* positive families. 51

5.4 Descriptive statistics for the probands of *BRCA2* positive families. 52

5.5 Descriptive statistics for the Mammographic Screening and Bilateral Oophorectomy ages in years 53

5.6 Modelling results for *BRCA1/2* mutation positive families with a shared frailty competing risk model with Permanent Exposure (PE) TVC for Bilateral Oophorectomy (BO) and Exponential Decay (ED) for Mammographic Screenings (MS). . 58

A.1 (1000 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Permanent Exposure TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations. 73

A.2 (1000 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Exponential Decay TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations. 74

A.3 (1000 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Cox and Oakes TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations. 75

A.4 (250 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Cox and Oakes TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations. 76

A.5 (250 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Cox and Oakes TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations. 77

A.6 (250 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Cox and Oakes TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations. 78

List of Abbreviations

AIC . . . . . Akaike Information Criterion

ASE . . . . . Average Standard Error

BC . . . . . Breast Cancer

BCFR . . . . . Breast Cancer Family Registry

BM . . . . . Bilateral Mastectomy

BO . . . . . Bilateral Oophorectomy

BRCA . . . . . BReast CAncer gene

CI . . . . . Confidence Interval

CO . . . . . Cox and Oakes

DNA . . . . . Deoxyribonucleic Acid

ECP . . . . . Empirical Coverage Probability

ED . . . . . Exponential Decay

ESE . . . . . Empirical Standard Error

HBOC . . . . . Hereditary Breast and Ovarian Cancer

HR . . . . . Hazard Ratio

KM . . . . . Kaplan Meier

MRI . . . . . Magnetic Resonance Imaging

MS . . . . . Mammographic Screening

OC . . . . . Ovarian Cancer

PE . . . . . Permanent Exposure

PS . . . . . Prophylactic Surgery

SD . . . . . Standard Deviation

SE . . . . . Standard Error

TIC . . . . . Time Invariant Covariate

TVC . . . . . Time Varying Covariate

Chapter 1

Introduction

In this thesis, we propose a shared frailty model extended to competing risks framework with time-varying covariates for data arising from family studies. We provide the cause-specific cumulative incidence function also called, penetrance, to estimate age-specific risks of disease with time-varying covariates in the presence of competing events. A shared frailty model is specified to account for familial residual dependence and an ascertainment-corrected likelihood is used to correct selection bias in family studies.

Our study is motivated by hereditary breast ovarian cancer (HBOC) families that experience high risks of breast and ovarian cancers due to mutations in *BRCA1/2* genes. *BRCA*1 and 2 are tumour suppressor genes that produce proteins that help repair damaged DNA, keeping the genetic material of the cell stable. A damaged *BRCA* gene can lead to increased risk of cancer, particularly breast or ovarian cancer in women and, to a lesser extent, other cancers such as prostate cancer in men. According to Petrucelli et al. (2016), for *BRCA1/2* mutation carriers, lifetime risk of breast cancer is from 38% to 87% while that of the general population is only about 12% for women. Ovarian cancer risk is also significantly increased for mutation carriers. Prevention and intervention such as mammographic screening (MS) and prophylactic surgery (PS) can potentially affect the cancer risks in *BRCA1/2* mutation positive families.

Our proposed modelling approach is aimed to handle the following statistical challenges arising from modeling HBOC family data: (1) the presence of competing risks, (2) familial correlation and ascertainment bias of sampling families and (3) time-dependent nature of covariates such as MS and PS.

First, while women in HBOC families are at risk for both breast and ovarian cancers while death from other causes than breast or ovarian cancer can preclude observing those cancers. Thus, we treat breast, ovarian cancers and death as competing events. The occurrence of ovarian cancer can alter the probability of breast cancer, or vice versa. For example, women affected by ovarian cancer could undergo various treatment regimes such as oophorectomy or radiotherapy treatment, which can potentially affect the risk of breast cancer. Therefore, modelling the risk of breast cancer is susceptible to bias when ignoring the presence of ovarian cancer in the risk set by treating the event as right censoring. In survival analysis, noninformative censoring is an important assumption which states that the mechanism behind the censoring must be independent of the survival time of the subject. In other words, the distribution of right censoring times bears no information on the distribution of the survival times. Subjects with censored survival times have the same survival prognosis as those who are still in the study at any given time, i.e., the censored subjects are at risk for the outcome regardless of the reason why they were censored for the duration of the study. However, this assumption is often violated in practice where the study population is exposed to other events referred to as competing events, which alter the probability of experiencing the event of interest. For instance, it is common practice to censor the patients who die before the occurence of the event of interest. These right censored subjects do not have the same prognosis as those who are still followed because subjects who die of other causes have a probability of zero to reach the event of interest. Without accounting for the competing risk of death, the incidence of the event will be overestimated (eg. Kaplan Meier estimates) since subjects censored due to the death remain in the risk set. This overestimation may be substantial with higher rate of occurence of the competing events (Noordzij et al., 2013).

Third, the breast and ovarian cancer are rare among the general population and highly hered-

itary, thus, family-based study is often designed to recruit higher risk individuals and their family members than general population. This invokes the two necessities in our modelling appoach: handling of the familial clustering among observations through frailty based model and the correction for the sampling bias through ascertainment correction. Family data possess an unobserved cluster effect which induces dependence among failure times in the same family. Frailty model is commonly used to model clustered data for which a random effect, referred to as frailty, is used to account for the variability in failure times within families. This random effect acts in family level to generate the dependence among family members and measures the effect of unobserved risk factors shared within the family.

Finally, most mutation carrier women opt for screening strategies rather than risk-reducing mastectomy with the hope that, if a cancer develops, it will be detected at a curable state. These screening strategies in women rely on a combination of monthly breast self-examination, clinical breast examination, mammography and breast MRI. Relative risk estimates of these treatments as well as prophylactic surgeries could help a woman making treatment decision e.g. chemo-prevention decision or the use of oral contraceptives. As those variables change over time, it is crucial to accurately incorporate them and evaluate their effects to minimize the misspecified exposure period. Because the subjects are often followed over a long time period, treating these variables as time invariant covariates might induce bias in the modelling result due to mispecified exposure time to the covariate effect. In this thesis, we suggest methodology to analyze their complex variation with time by incorporating the Permanent Exposure (PE), Exponential Decay (ED) and Cox and Oakes (CO) functions for the effect of binary time varying covariates.

1.1 Objectives

The objectives of our study are as follow:

1. Develop shared frailty competing risk models with time varying covariates to account for the ascertainment bias and familial correlation;

2. Evaluate and compare the performances of the three time varying covariate functions in shared frailty competing risks models through a simulation study;

3. Derive the cumulative incidence function for an event of interest in the presence of time varying covariates using the proposed shared frailty competing risk model;

4. Assess performances of the proposed models on the estimation of cumulative incidence risk through a simulation study;

5. Estimate the risk of the first cancer (either breast or ovarian) in *BRCA1/2* mutation positive families accounting for a woman's screening history and the prophylactic surgery treating them as time varying covariates using the proposed model.

1.2 Organization of the thesis

In Chapter 2, we provide a literature reviews on related topics including competing risk models, the shared frailty model for clustered data, time varying covariates and ascertainment bias. In Chapter 3 we develop a shared frailty competing risks model to account for three different effects of time varying covariates and provide cause-specific penetrance accounting for time-dependent covariates. In Chapter 4, we conduct a simulation study to evaluate our proposed model for estimating the penetrances. In Chapter 5, we apply the proposed model to data from *BRCA1* and *BRCA2* mutation positive families collected by the Breast Cancer Family Registry. Possible future research and limitations are presented in Chapter 6.

Chapter 2

Literature Review

2.1 Competing risk models

Competing risk models are designed to extend the traditional survival models by considering the collection of mutually exclusive events. In some clinical settings, subjects are exposed to more than one disease, which can potentially influence the occurrence of the disease of interest. For instance, when analyzing peritoneal dialysis patients, the time to peritonitis episode is of interest. However, before this episode, the patient might undergo death, kidney transplantation or transfer to heamodialysis. These events are then competing for its first occurrence where one occurrence can affect the probabilities of the others (Noordzij et al., 2013). In this scenario, one might completely ignore competing events or simply right censor the subjects who are affected by them. The latter approach is not to promising since it potentially violates the non-informative censoring assumption, which states that censoring mechanism must be independent of the survival time of the subject. Then, the violation of the assumption induces the bias in the risk estimation. As an example, suppose we right censor the patients who died before the occurence of the peritonitis episode, but assume these patients have the same prognosis as the subjects still in the study. However, deceased subjects cannot be affected by peritonitis, which means even though they contribute to the risk set, their probability of developing the episode is 0 leading to the overestimation of its risk (e.g. Kaplan Meier estimates). According

to Koller et al. (2012), 70% of the 50 articles in top journals neglected competing risk process or imposed naive-approach even while their study populations were susceptible to competing risks.

Methodologies to handle competing risks have been developed by several authors. Prentice et al. (1978) proposed cause-specific hazard approach that modeled the hazard function separately for each type of failure. The subdistribution hazard regression model introduced by Fine and Gray (1999) has its significance in terms of one-to-one correspondence between its hazard function for the specific event and the cumulative incidence of the same event, whereas it is not viable in cause-specific approach. The mixture model by Larson and Dinse (1985) and Ng and McLachlan (2003) involves fitting a more complex model which assumes multinomial distribution for failure types. Their model expresses the failure time distribution in competing risk data in terms of the marginal distribution of failutre type (multinomial logistic distribution) and the conditional distribution of time to failure, given the type of failure (proportional hazard model). This mixture model assumes that an individual will fail from a particular risk, chosen by a stochastic mechanism characterized by the marginal distribution of each failure type (Ng and McLachlan, 2003). Lune and McNeil (1995) modified Cox regression to treat the different types of failure jointly by augmenting the data using a duplication method. With J failure types, the data would be duplicated J times, one row for each failure type. Each event type is defined as an indicator variable that equals to 1 for event j and 0 for the remaining $J-1$ entries. An unstratified or a stratified Cox regression could be applied on this augmented dataset, depending on whether the assumption of proportional hazards holds. The unstratified method assumes that the hazard functions for different failure types are proportional, whereas the stratified method allows for different baseline hazards in each failure type. The primary advantage of this procedure is that it facilitates direct comparisons between different event types.

This thesis is based on the cause-specific approach due to the ease of incorporating cause-specific frailty and TVCs. The cause-specfic hazard approach was introduced by Prentice et al. (1978). Cause-specific hazard function, $h_j(t)$, is defined as the instantaneous rate of occurrence of the specific event, within a very short period of time, t and $t+dt$, given that the subject has

survived by time t from all the competing events until time t. Let T^o and C be the time to the first event time and the right censoring time, respectively. Let $\delta \in \{1,\ldots,J\}$ be the type of the first observed event time among J competing events and $\delta = 0$ if right censored. Define $T = \min(T^o, C)$. Then, the cause-specific hazard for event j, $j = 1,...,J$, is defined as,

$$h_j(t) = \lim_{dt \to 0} \frac{1}{dt} P(t \leq T < t + dt, \delta = j | T \geq t).$$

The corresponding cause-specific cumulative hazard, $H_j(t)$ is obtained by integrating the cause-specific hazard until time t, $H_j(t) = \int_0^t h_j(u)du$. Then, the probability of surviving from all the possible events until time t is defined as

$$S(t) = P(T \geq t) = \exp\Big\{ - \sum_{j=1}^{J} \int_0^t h_j(u)du \Big\}.$$

Finally, the cause-specific cumulative incidence for event j is defined as

$$F_j(t) = P(T \leq t, \delta = j) = \int_0^t h_j(u) S(u) du.$$

The idea behind this relation can be explained as follows. The probability of the given event j occurring at any time u is determined by the product of the two terms inside the integral: One is the probability that the subject has not experienced any event before time u indicated by overall survival function $S(u) = P(T \geq u)$ and the other is the probability of event j occurring at time u given that no event has happened. Since the cause-specific cumulative incidence function associates all the possible causes in its formulation, there is no one-to-one correspondence between cause-specific hazard and its cause-specific cumulative incidence (Geskus, 2016). For this reason, the way that covariates are associated with cause-specific hazard and cause-specific cumulative incidence may not coincide in the regression setting.

Nonparametric estimation of the cause specific cumulative incidence can be obtained by the method originally aimed for estimating transition probability of multi-state model introduced by Aalen and Johansen (1978) as

$$\hat{F}_j^{AJ}(t) = \sum_{Z_i \leq t} \hat{S}(Z_i)\hat{h}_j(Z_i),$$

where Z_i is the ith event time among the uniquely ordered M event times, $\{Z_1, Z_2, ..., Z_M\}$ and $\hat{h}_j(Z_i)$ is the cause-specific hazard at each observed discrete time Z_i and obtained by the number of subjects with event j at Z_i, divided by the total number of subjects at risk. $\hat{S}(Z_i)$ is the classical Kaplan-Meier estimate of the overall survival obtained by simply combining all the event types.

Southern et al. (2006) and Tai et al. (2001) compared the naive Kaplan Meier (KM) estimators under three censoring approaches in the presence of competing risks: 'censor all', 'censor death only', and 'ignore all'. In the 'censor all' approach, subjects are censored with the occurrence of a competing event in advance of the event of interest. In the 'censor death only' approach, subjects are censored if they experience death before the event of interst. Lastly, the 'ignore all' approach only censors subjects who are truly lost to follow-up. Upward bias in cumulative incidence of event of interest was found to be the most significant in 'censor all' method followed by 'censor death only' and 'ignore all' methods. It was clearly shown that all of the naive KM estimators with different censoring approaches overestimated the cause-specific cumulative incidence. Additionally, Tai et al. (2001) showed that estimates of the cause-specific cumulative incidence from the cause-specific hazard model and the modified Cox method by Lune and McNeil (1995) were highly comparable.

Shifting our focus on the regression model to incorporate the covariate effects, suppose we have a vector of covariates $\boldsymbol{X}$. The cause-specfic hazard model for event j given $\boldsymbol{X}$ is defined as

$$h_j(t|\boldsymbol{X}) = h_{0j}(t)\exp\{\boldsymbol{\beta}_j^T\boldsymbol{X}\},$$

where $\boldsymbol{\beta}_j$ is the vector of regression coefficients associated with event j and $h_{0j}(t)$ is the baseline hazard function, which can be modeled parametrically in specific functional form. Several common baseline hazard models are specified in Table 2.1. We can also subdivide time into reasonably small intervals and assume that the baseline hazard is constant in each interval,

leading to a piece-wise semi-parametric model. A baseline hazard function needs not to be specified if we rely on the partial likelihood function on model fitting focusing only on the estimation of regression parameters.

No additional assumption on the dependence among competing events is required to estimate cause-specific hazard (Prentice et al., 1978, Kalbfleisch and Prentice, 2002). Standard techniques to fit traditional Cox proportional hazard model can be used to fit the cause-specific hazard model for each competing event. The covariate effect has an interpretation on the cause-specific hazard.

Table 2.1: Several possible choices of baseline hazard function

Distribution	hazard $h_{0j}(t)$	cumulative hazard $H_{0j}(t)$
Weibull	$\rho_j\lambda_j(\lambda_j t)^{\rho_j-1}$	$(\lambda_j t)^{\rho_j}$
Log-logistic	$\frac{\rho_j\lambda_j(\lambda_j t)^{\rho_j-1}}{1+(\lambda_j t)^{\rho_j}}$	$\log\{1+(\lambda_j t)^{\rho_j}\}$
Gompertz	$\lambda_j e^{\rho_j t}$	$\frac{\lambda_j}{\rho_j}(e^{\rho_j t}-1)$

2.2 Frailty models for clustered time-to-event data

Clustered time-to-event data often arise in clinical research. For instance, in multicenter clinical trials, patients in each center could potentially share common factors exclusive to their center, so their clinical outcomes could be similar. Another example arises from family studies where the ages at onset of disease are correlated among family members. Multiple records observed from the same individual can be also viewed as clustered failure times. It is well noted that inferences ignoring this cluster effect are potentially misleading; a misspecified model can lead to underestimation of the covariate coefficients (Keilding et al., 1997). Lee et al. (1992) have shown that it is still possible to obtain consistent and asymptotically normal estimators of covariate coefficients in the standard Cox regression ignoring cluster effect. However, corresponding variance-covariance estimators is not valid due to the dependence among members in the same cluster. Approaches to obtain valid variance-covariance estimators accounting for cluster effect are discussed by Lee et al. (1992) and Spiekerman and Lin (1998). How-

ever, when the interest lies in obtaining the cluster effect, dependence structure between failure times must be modelled. Shared frailty models and copula models are popular choices in this situation. Similarities and differences between frailty and copula models were discussed by Goethals et al. (2008). Considering the case of two subjects in each cluster, the joint survival function of the two failure times is modeled via a function, called copula, which connects the marginal survival functions of the two failure times. On the other hand, the frailty based model shares a random variable, also called frailty, within a cluster. The frailty variable has a multiplicative effect on the hazard function. Conditional on the frailty term, failure times within a cluster are assumed to be independent. Then, a joint survival function can be obtained by integrating out the frailty.

2.2.1 Shared frailty model in non-competing risks setting

The frailty model was introduced by Vaupel et al. (1979) in survival model framework to account for the unobserved heterogeneity in study population. Furthermore, Clayton (1978), Hougaard (1984) and Oakes (1982) extended the frailty model to account for the unobserved heterogenity shared within clusters introducing parametric gamma frailty distribution.

Let $\boldsymbol{X}_{f_i}$ be the vector of covariates for the subject i in cluster f and z_f be the shared frailty for cluster f. This shared frailty acts as a multiplicative factor on the the hazard under Cox regression model as

$$h_{f_i}(t|\boldsymbol{X}_{f_i}, z_f) = h_0(t) z_f \exp(\boldsymbol{\beta}^T \boldsymbol{X}_{f_i}),$$

where $h_0(t)$ is unspecified baseline hazard. Frailties z_f are assumed to be independently and identically distributed random variables. For the distribution of frailty, log-normal, gamma, or inverse Gaussian are the popular choices (Duchateau and Janssen, 2008). We continue our discussion with one parameter gamma distribution proposed by Hougaard (2000) with shape parameter k and scale parameter $1/k$, $z_f \sim Gamma(k, 1/k)$ to take an advantage of its mathematical simplicity. Its density function follows

$$g(z_f) = \frac{z_f^{k-1} e^{-k z_f}}{k^{-k}\Gamma(k).}$$

Under this specification, random variable z_f has mean of 1 and the variance of $1/k$. Thus, the smaller value of k induces the stronger within-cluster dependence. Since we do not observe the frailty variable in practice, we obtain marginal distribution of the failure times by integrating out z_f in the conditional distribution using the Laplace transform of the frailty distribution. When $k \to \infty$, it is equivalent to the independent model and the hypothesis test $H_0 : 1/k = 0$ can be constructed using Likelihood Ratio (LR) statistics. In this case, the null hypothesis $1/k = 0$ is on the boundary of the parameter space and the distribution of LR test-statistics has a 50:50 mixture of a chi-square with no degrees of freedom and a chi-square with one degree of freedom (Self and Liang, 1987). $k \to \infty$ means $z_f = 1$ and frailty effects are not important. Kendall's τ, which is a non-parametric measure of correlation, is related to the frailty parameter k such that $\tau = \frac{1}{1+2k}$ (Munda et al., 2012). Values of τ close to 1 indicate higher dependence between failure times within clusters.

2.2.2 Shared frailty model in competing risk setting

So far our discussion was restricted to the clustered survival data under the non-competing risk setting that accounts for the dependence of failure times of the one event type within a cluster. In competing risk setting, dependence of the failure times could be different for different competing events. Gorfine and Hsu (2011) and Lai et al. (2017) discussed the methods for modelling the dependence structure between the random frailty variables assigned for each event type. Restricting our attention to the methodologies relevant to the cause-specific hazard model framework, Gorfine and Hsu (2011) proposed multivariate normal distribution between frailty variables.

We now denote $\boldsymbol{z}_f = \{z_{f_1}, ..., z_{f_J}\}$ as the vector of random frailty variables in cluster f for J competing events. Gorfine and Hsu (2011) proposed a competing risk model where cause-specific hazard follows

$$h_{f_i j}(t|\boldsymbol{X}_{f_i}, z_{f_j}) = \lim_{dt \to 0} \frac{1}{dt} P(t \leq T_{f_i} < t + dt, \delta_{f_i} = j | T_{f_i} \geq t, \boldsymbol{X}_{f_i}, z_{f_j})$$
$$= h_{0j}(t) \exp(\boldsymbol{\beta}_j^T \boldsymbol{X}_{f_i} + z_{f_j}),$$

for $j = 1, ..., J$, where $\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_J$ are the vectors of cause-specific regression coefficients and $h_{01}(\cdot), ..., h_{0J}(\cdot)$ are cause-specific baseline hazard functions. Importantly, Gorfine specified the dependence structure of the $\boldsymbol{z}_f$ with

$$\boldsymbol{z}_f = \{z_{f_1}, ..., z_{f_J}\} \sim N_J(\boldsymbol{\mu}, \Sigma),$$

where $N_J(\boldsymbol{\mu}, \Sigma)$ is the multivariate normal distribution with J-dimensional mean vector $\boldsymbol{\mu}$ and $J \times J$ covariance matrix. Estimation of Σ gives insight to the dependence of failure times of J competing events between subjects in the same cluster. For instance, in case of bivariate normal distribution accounting for two competing events, Σ can be defined as:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

where the correlation coefficient ρ_{12} measures the dependence between the frailty terms z_{f_1} and z_{f_2}. If $\rho_{12} = 0$, the model degererates to the model with independent frailties. In other words, z_{f_j}, $j = 1, ..., J$, are independent between different events. Lai et al. (2017) proposed a bivariate normal random effects competing risk model under two competing events. It is similar to Gorfine's model but they used iterative numerical procedure to estimate the model parameters under generalized linear mixed model framework. In the first step, they obtain estimates of covariate coefficients and conditionally fixed random cluster effects for a given variance components parameters by maximizing partial log-likelihood. In the second step, they obtained restricted maximum likeliehood estimates of variance components parameters in bivariate normal distribution by solving the estimating equations.

Lee et al. (2017) discussed similar approach but varying degree of dependence structure.

Lee et al. (2017) assumed each cluster f shares the same frailty variable z_f for all the competing events. Under this specification, jth cause-specific hazard becomes

$$h_{f_i j}(t|\boldsymbol{X}_{f_i}, z_f) = h_{0j}(t) z_f \exp\{\boldsymbol{\beta}_j^T \boldsymbol{X}_{f_i}\}.$$

In this case, z_f accounts for the unobserved cluster-level effect not specific to event j but all the possible events. They also extended this framework to handle missing cause of failure using hierarchical likelihood.

2.3 Time varying covariates

In analyzing time to event data, Cox proportional hazard model has the advantage of ease of incorporation on the effect of covariates as multiplicative factor to the unspecified baseline hazard. Time-fixed or time invariant covariate (TIC), which is measured at baseline, does not change over the study period. However, in practice, it is plausible to assume that covariate values change over time due to, for instance, drug dosages or the number of screenings which vary over follow-up period. Cox model allows us to incorporate such time varying covariates (TVC) denoted as $X(t)$ based on the modelling of hazard function $h(t)$. Hazard function describes the instantaneous risk of the event of the interest at time t given that the subject has survived until that time. As we condition on the survival until time t, covariate value can be measured at any time between 0 and t.

In clinical setting, assessing the effect of TVC is important especially when the follow-up duration is long. For example, consider binary variable for certain treatment occurring later period of the follow-up duration. If we code this variable as TIC, treatment exposure duration becomes much longer than actual exposure. We lose information that the subject was actually absent of its effect for the most of the follow-up time. Therefore, modelling results can be potentially misleading due to the bias incurred. Daniel et al. (2015) quantified bias in Cox regression for the study of risk of spontaneous abortion following nonsteroidal anti-inflammatory drug exposure. When the exposure of the drug was treated as TIC, the drug effect was statis-

tically significant but this effect was not significant under TVC setting. In addition, the effects of all the other drugs in this study were biased in varying extent depending on the length of misspecified exposure period. Our analysis on the effect of the BO and MS in Chapter 5 is relevant to this issue since most of the patients have relatively long follow-up duration.

Suppose subjects received treatment at time t_x during the follow-up. The TVC can be defined as $X(t,t_x) = 0$ for $t < t_x$ and $X(t) = 1$ for $t \geq t_x$. Then, the effect of treatment in Cox model is multiplicative to the baseline hazard such that $h_0(t)e^{\beta X(t,t_x)}$. This type of TVC is referred to as Permanent Exposure (PE) (Keown-Stoneman et al., 2018) as its effect stays constant at β as $t \to \infty$. The effect of the PE does not change over time t since the treatment time t_x. In other words, the hazard of a subject who received treatment recently does not differ from that of a subject who received the treatment a long time in the past. This assumption may not be plausible in many real world applications. One of the possible resolutions to this problem is suggested by Cox and Oakes (1984). They proposed the formulation of a TVC effect which decays over time with the rate parameter η and the additional parameter η_0 that measures the converged effect of TVC. Under this setting, the corresponding hazard function follows as

$$h(t|\boldsymbol{\theta}) = \begin{cases} h_0(t) & \text{if } t < t_x \\ h_0(t)\exp\{\eta_0 + \beta e^{-\eta(t-t_x)}\} & \text{if } t \geq t_x. \end{cases}$$

In this formulation, the effect of the TVC on baseline hazard decreases over time with the rate of η and converges to η_0 as $t \to \infty$. Subjects who received the treatment recently are under stronger effect of the treatment than those who received the treatment further in the past. Cox and Oakes noted multiple local maxima may exist in the partial likelihood function. As a simplication of the Cox and Oakes model with $\eta_0 = 0$, Keown-Stoneman et al. (2018) suggested Exponential Decay (ED) model as

$$h(t|\boldsymbol{\theta}) = \begin{cases} h_0(t) & \text{if } t < t_x \\ h_0(t)\exp\{\beta e^{-\eta(t-t_x)}\} & \text{if } t \geq t_x. \end{cases}$$

Under this model, the effect of TVC converges to 0 as $t \to \infty$. They also provided the

method to generate data under this model and performed a hypothesis test for $H_0 : \eta = 0$ against $H_1 : \eta \geq 0$ using likelihood ratio test. When $\eta = 0$, this model is equivalent to the permanent exposure model. Figure 2.1 graphically displays the effects of three TVC models.

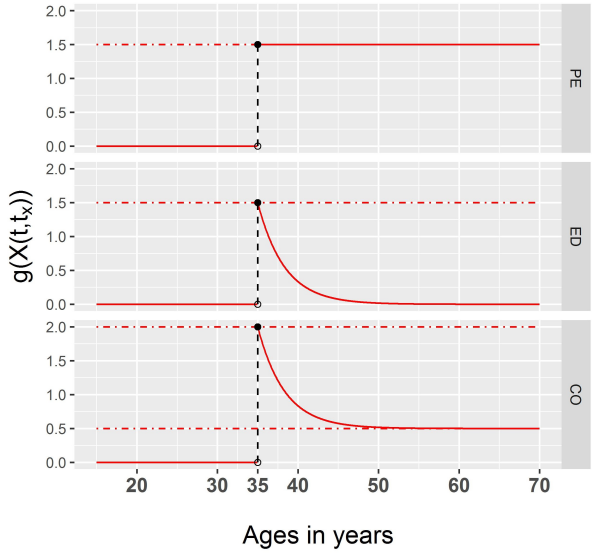

Figure 2.1: Effect of time-varying covariate over time in Permanent Exposure (PE), Exponential Decay (ED) and Cox and Oakes (CO) models given $t_x = 35$, $\beta = 1.5$, $\eta_0 = 0.5$ and $\eta_1 = 0.3$.

2.4 Ascertainment bias and ascertainment corrected likelihood

Ascertainment bias in family studies arises due to the way individuals were selected. They were only recruited via having high risk scores depending on familial disease history or mutation carrier status in genes related to the disease. Therefore, collected individuals do not represent the general population. Ignoring this selection procedure will incur the bias on the estimation of the model parameters such as relative risk for the mutation carrier and non-carrier and penetrance function. Gong and Whittemore (2003) discussed population and clinic based study designs which are distinguished by their inclusion criteria. In population based design, families are ascertained via an individual identified as affected or mutation carrier, where this

individual is referred to as the proband. In clinic based design, multiple family members must be affected for the study entry. Choi et al. (2008) developed the ascertainment corrected likelihood methods for population- and clinic-based family designs and evaluated performances of study designs on the estimation of penetrance and relative risk via a simulation study.

Suppose we have n independent families and family f, $f = 1, ..., n$ has n_f members. Ascertainment corrected likelihood has a general form:

$$L_C = \prod_{f=1}^{n} \frac{L_f}{A_f},$$

where L_f is the likelihood contribution of family f and A_f is the probability of the family being ascertained.

Chapter 3

Statistical Models

In this chapter, we develop the shared frailty competing risk model with Time Varying Covariates (TVC) and explain the inference procedure. In Section 3.1, we construct shared frailty competing risk model with TVC. In Section 3.2, we speculate about likelihood construction with ascertainment correction. In Section 3.3, we derive cause-specific penetrance function with TVC. We discuss variance estimation procedure for cause-specific penetrance and covariate coefficients in Section 3.4.

3.1 Shared frailty competing risk model with time-varying covariates

Consider n independent families, with family f, $f = 1,\ldots,n$, having n_f members. For family member i of the family f, let $T^o_{f_i}$ and C_{f_i} be the time to the first event time and the right censoring time, respectively. Let $\delta_{f_i} \in \{1,\ldots,J\}$ be the type of the first observed event time among J competing events and $\delta_{f_i} = 0$ if right censored. Define $T_{f_i} = \min(T^o_{f_i}, C_{f_i})$. Let $\boldsymbol{X}_{f_i}$ be the vector of covariates which consist of time invariant and time varying covariates for individual i in family f. Finally, we denote z_{f_j} as the shared frailty specific to the event j within family f. Conditional on the vector of covariates $\boldsymbol{X}_{f_i}$ and the cause-specific familial frailty z_{f_j}, we assume the cause-specific hazard function for event j for individual i from family f follows

a proportional hazards regression model

$$\begin{aligned} h_{f_i j}(t|\boldsymbol{X}_{f_i}, z_{f_j}) &= \lim_{dt\to 0} \frac{1}{dt} P(t \le T_{f_i} < t+dt, \delta_{f_i} = j | T_{f_i} \ge t, \boldsymbol{X}_{f_i}, z_{f_j}) \\ &= h_{0j}(t) z_{f_j} e^{\boldsymbol{\beta}_j^T \boldsymbol{X}_{f_i}}, \end{aligned} \tag{3.1}$$

where h_{0j} is the baseline hazard function and $\boldsymbol{\beta}_j$ is the vector of the covariate effects related to event j. The corresponding jth cause-specific cumulative hazard function is defined as

$$\begin{aligned} H_{f_i j}(t|\boldsymbol{X}_{f_i}, z_{f_j}) &= \int_0^t h_{f_i j}(u|\boldsymbol{X}_{f_i}, z_{f_j}) du \\ &= H_{0j}(t) z_{f_j} e^{\boldsymbol{\beta}_j^T \boldsymbol{X}_{f_i}}, \end{aligned} \tag{3.2}$$

where H_{0j} is the cumulative hazard function for event j.

By the standard theory of competing risks, given the covariates and frailty, we define the conditional cumulative hazard function H and survival function S:

$$\begin{aligned} H_{f_i}(t|\boldsymbol{X}_{f_i}, z_{f_j}) &= \sum_{j=1}^{J} H_{f_i j}(t|\boldsymbol{X}_{f_i}, z_{f_j}) \\ S_{f_i}(t|\boldsymbol{X}_{f_i}, z_{f_j}) &= \exp\{-\int_0^t \sum_{j=1}^{J} h_{f_i j}(u|\boldsymbol{X}_{f_i}, z_{f_j}) du\} \\ &= \exp\{-H_{f_i}(t|\boldsymbol{X}_{f_i}, z_{f_j})\}, \end{aligned} \tag{3.3}$$

where these are overall quantities across all competing events.

We implement time varying covariates (TVC) into the shared frailty competing risks model in equations (3.1) to (3.3). For some covariates, such as individual *BRCA1/2* mutation status, they are time independent. However, for some covariates, such as surgery or screening variables which occur at certain age, it is plausible to assume that its effect might change over time. Without loss of generality, suppose $\boldsymbol{X}_{f_i}$ includes a time varying covariate $X_{f_i}(t, t_x)$ and a time invarying covariate G_{f_i}. We assume that $X_{f_i}(t, t_x) = 0$ at $t < t_x$ and 1 at $t \ge t_x$, where t_x is the time that change in value of time varying covariate occured. Then, the effect of the TVC that changes over time, denoted by $g(\cdot)$, can be described in three different structures: PE, ED, and CO as follows,

$$g(X_{f_i}(t)) = \begin{cases} 0 & \text{if } t < t_x \text{ (PE,ED,CO)} \\ \beta & \text{if } t \geq t_x \text{ (PE)} \\ \beta \exp\{-\eta(t-t_x)\} & \text{if } t \geq t_x \text{ (ED)} \\ \beta \exp\{-\eta(t-t_x)\} + \eta_0 & \text{if } t \geq t_x \text{ (CO)} , \end{cases}$$

where for time $t \geq t_x$, the effect of TVC stays at β for PE, whereas it starts to decrease with a rate of η to 0 for ED or to η_0 for CO. Then, we can specify jth cause-specific hazard and cumulative hazard function with TVC and TIC as

$$h_{f_i j}(t|\boldsymbol{X}_{f_i}, z_{f_j}) = h_{0j}(t) z_{f_j} \exp\{g(X_{f_i}(t,t_x)) + \alpha_j G_{f_i}\}.$$

$$H_{f_i j}(t|\boldsymbol{X}_{f_i}, z_{f_j}) = \int_0^t h_{0j}(u) z_{f_j} \exp\{g(X_{f_i}(u,t_x)) + \alpha_j G_{f_i}\} du,$$

where G_{f_i} is the TIC and α_j is the corresponding cause-specific covariate coefficient for event j. For simplicity, omitting the terms related to TIC and the shared frailty term, the cause-specific hazard can be specified for PE, ED and CO as

$$h_{f_i j}(t|\boldsymbol{X}_{f_i}) = \begin{cases} h_{0j}(t) & \text{if } t < t_x \text{ (PE,ED,CO)} \\ h_{0j}(t) \exp\{\beta_j\} & \text{if } t \geq t_x \text{ (PE)} \\ h_{0j}(t) \exp\{\beta_j e^{-\eta_j (t-t_x)}\} & \text{if } t \geq t_x \text{ (ED)} \\ h_{0j}(t) \exp\{\beta_j e^{-\eta_j (t-t_x)} + \eta_{0_j}\} & \text{if } t \geq t_x \text{ (CO)} . \end{cases}$$

$$H_{f_i j}(t|\boldsymbol{X}_{f_i}) = \begin{cases} H_{0j}(t) & \text{if } t < t_x \text{ (PE,ED,CO)} \\ H_{0j}(t_x) + \big(H_{0j}(t) - H_{0j}(t_x)\big) \exp\{\beta_j\} & \text{if } t \geq t_x \text{ (PE)} \\ H_{0j}(t_x) + \int_{t_x}^t h_{0j}(u) \exp\{\beta_j e^{-\eta_j (u-t_x)}\} du & \text{if } t \geq t_x \text{ (ED)} \\ H_{0j}(t_x) + \int_{t_x}^t h_{0j}(u) \exp\{\beta_j e^{-\eta_j (u-t_x)} + \eta_{0_j}\} du & \text{if } t \geq t_x \text{ (CO)} . \end{cases}$$

where β_j,η_j and η_{0j} are the cause-specific TVC effect parameters for event j. Numerical integration is required for computing cumulative hazard for ED and CO since no closed form exists.

3.2 Likelihood construction with ascertainment correction

Let $\boldsymbol{\theta} = \{h_{0j}(.), \boldsymbol{\beta}_j, k_j, j = 1,\ldots,J\}$ be the vector of parameters involved in the model, which consists of baseline parameters for specifying baseline hazard functions, regression coefficient vector $\boldsymbol{\beta}_j$ and frailty parameter k_j for each competing event $j = 1,...,J$. Then, the likelihood of the data from n families can be constructed simply by the product of the likelihoods of all families.

$$L(\boldsymbol{\theta}) = \prod_{f=1}^{n} L_f(\boldsymbol{\theta}).$$

Under the shared frailty competing risk model framework, the likelihood for family f is obtained by integrating over the frailty distribution:

$$L_f(\boldsymbol{\theta}) = \prod_{i=1}^{n_f} \underbrace{\int_0^\infty \cdots \int_0^\infty}_{J} \{\prod_{j=1}^{J} h_{f_i j}(t_{f_i}|\boldsymbol{X}_{f_i}, z_{f_j})^{\mathrm{I}(\delta_{f_i}=j)}\} S_{f_i}(t_{f_i}|\boldsymbol{X}_{f_i}, \boldsymbol{z}_f) g(z_{f_1}) \ldots g(z_{f_J}) dz_{f_1} \ldots dz_{f_J}. \tag{3.4}$$

To compute the integrals, we utilize Laplace transform $\phi(\cdot)$ of the frailty distribution $g(z_{f_j})$ and its dth derivative, $\phi(\cdot)^{(d)}$, which have the following expressions

$$\begin{aligned} \phi(s) &= \int_0^\infty e^{-sz} g(z) dz \\ \phi(s)^{(d)} &= (-1)^d \int_0^\infty z^d e^{-sz} g(z) dz. \end{aligned} \tag{3.5}$$

With Gamma frailty distribution, they have closed form expressions:

$$\begin{aligned} \phi(s) &= \left(1 + \frac{s}{k}\right)^{-k} \\ \phi(s)^{(d)} &= (-1)^d \frac{(k+d-1)!}{k! k^{d-1}} \left(1 + \frac{s}{k}\right)^{-k-d}. \end{aligned} \tag{3.6}$$

With the simple case of only 2 competing events, $J = 2$, which can be easily generalized to

any number of J, the likelihood for family f can be obtained, omitting $\boldsymbol{X}_{f_i}$ after the first line for simplicity:

$$
\begin{aligned}
L_f(\boldsymbol{\theta}) &= \int_0^\infty \int_0^\infty \prod_{i=1}^{n_f} \Big\{ \prod_{j=1}^{2} h_{f_i j}(t_{f_i}|\boldsymbol{X}_{f_i}, z_{f_j})^{\mathrm{I}(\delta_{f_i}=j)} \Big\} \exp\Big\{ -\sum_{j=1}^{2} H_{f_i j}(t_{f_i}|\boldsymbol{X}_{f_i}, z_{f_j}) \Big\} g(z_{f_1}) g(z_{f_2}) dz_{f_1} dz_{f_2} \\
&= \int_0^\infty \int_0^\infty \prod_{i=1}^{n_f} \Big\{ \prod_{j=1}^{2} \big(z_{f_j} h_{ij}(t_{f_i})\big)^{\mathrm{I}(\delta_{f_i}=j)} \Big\} \exp\Big\{ -\sum_{j=1}^{2} z_{f_j} H_{ij}(t_{f_i}) \Big\} g(z_{f_1}) g(z_{f_2}) dz_{f_1} dz_{f_2} \\
&= \prod_{i=1}^{n_f} \prod_{j=1}^{2} h_{ij}(t_{f_i})^{\mathrm{I}(\delta_{f_i}=j)} \iint z_{f_1}^{d_{f_1}} z_{f_2}^{d_{f_2}} \exp\Big\{ -z_{f_1} \sum_{i=1}^{n_f} H_{i1}(t_{f_i}) - z_{f_2} \sum_{i=1}^{n_f} H_{i1}(t_{f_i}) \Big\} g(z_{f_1}) g(z_{f_2}) dz_{f_1} dz_{f_2} \\
&= \prod_{i=1}^{n_f} \prod_{j=1}^{2} h_{ij}(t_{f_i})^{\mathrm{I}(\delta_{f_i}=j)} \int z_{f_2}^{d_{f_2}} \exp\big\{ -z_{f_2} \sum_{i=1}^{n_f} H_{i2}(t_{f_i}) \big\} \int z_{f_1}^{d_{f_1}} \exp\big\{ -z_{f_1} \sum_{i=1}^{n_f} H_{i1}(t_{f_i}) \big\} g(z_{f_1}) dz_{f_1} g(z_{f_2}) dz_{f_2} \\
&= \prod_{i=1}^{n_f} \prod_{j=1}^{2} h_{ij}(t_{f_i})^{\mathrm{I}(\delta_{f_i}=j)} (-1)^{-d_{f_1}} \phi^{(d_{f_1})} \Big(\sum_{i=1}^{n_f} H_{i1}(t_{f_i}) \Big) \int z_{f_2}^{d_{f_2}} \exp\big\{ -z_{f_2} \sum_{i=1}^{n_f} H_{i2}(t_{f_i}) \big\} g(z_{f_2}) dz_{f_2} \\
&= \prod_{i=1}^{n_f} \prod_{j=1}^{2} h_{ij}(t_{f_i})^{\mathrm{I}(\delta_{f_i}=j)} (-1)^{-d_{f_1}} \phi^{(d_{f_1})} \Big(\sum_{i=1}^{n_f} H_{i1}(t_{f_i}) \Big) (-1)^{-d_{f_2}} \phi^{(d_{f_2})} \Big(\sum_{i=1}^{n_f} H_{i2}(t_{f_i}) \Big) \\
&= \prod_{i=1}^{n_f} \prod_{j=1}^{2} h_{ij}(t_{f_i})^{\mathrm{I}(\delta_{f_i}=j)} \prod_{j=1}^{2} \frac{(k_j + d_{f_j} - 1)!}{k_j! \cdot k_j^{d_{f_j}-1}} \Big(1 + \frac{\sum_{i=1}^{n_f} H_{ij}(t_{f_i})}{k_j} \Big)^{-k_j - d_{f_j}},
\end{aligned}
\tag{3.7}
$$

where d_{f_j} is the number of family members affected by event j. Furthermore, to correct for the ascertainment bias, we implement ascertainment corrected likelihood approach (Choi et al., 2008). Families are ascertained via the probands (indexed as p) who are affected by their age at examination (a_{f_p}). For each family f, we divide the $L_f(\boldsymbol{\theta})$ by the probability of the proband being ascertained by her age at examination. Then, this probability is the conditional cumulative incidence of events that qualify for ascertainment by a_{f_p}, that is $A_f(\boldsymbol{\theta}) = P(T_{f_p} \leq a_{f_p}|\boldsymbol{X}_{f_p})$. Therefore, ascertainment corrected likelihood of data follows

$$
L_C(\boldsymbol{\theta}) = \prod_{f=1}^{n} \frac{L_f(\boldsymbol{\theta})}{A_f(\boldsymbol{\theta}).}
$$

Recalling equations (3.3), (3.5) and (3.6), we derive this ascertainment correction term as,

$$
\begin{aligned}
A_f(\boldsymbol{\theta}) &= 1 - S_{f_p}(a_{f_p}|\boldsymbol{X}_{f_p}) \\
&= 1 - \int \cdots \int_0^\infty S_{f_p}(a_{f_p}|\boldsymbol{X}_{f_p},\boldsymbol{z}_f) g(z_{f_1}) \ldots g(z_{f_J}) dz_{f_1} \ldots dz_{f_J} \\
&= 1 - \int \cdots \int_0^\infty \exp\Big(- \sum_{j=1}^{J} z_{f_j} H_{f_p j}(a_{f_p}|\boldsymbol{X}_{f_p}) \Big) g(z_{f_1}) \ldots g(z_{f_J}) dz_{f_1} \ldots dz_{f_J} \\
&= 1 - \prod_{j=1}^{J} \phi\left(H_{f_p j}(a_{f_p}|\boldsymbol{X}_{f_p}) \right) \\
&= 1 - \prod_{j=1}^{J} \left(1 + \frac{H_{f_p j}(a_{f_p}|\boldsymbol{X}_{f_p})}{k_j} \right)^{-k_j}.
\end{aligned}
\tag{3.8}
$$

Then combining results from (3.7) and (3.8), we specify the ascertainment corrected likelihood for all the families:

$$
L_C(\boldsymbol{\theta}) = \prod_{f=1}^{n} \frac{\prod_{i=1}^{n_f} \prod_{j=1}^{J} \left\{ h_{f_i j}(t_{f_i}|\boldsymbol{X}_{f_i}) \right\}^{\mathrm{I}(\delta_{f_i}=j)} \times \prod_{j=1}^{J} \frac{(k_j+d_{f_j}-1)!}{k_j! k_j^{d_{f_j}-1}} \left(1 + \frac{\sum_{i=1}^{n_f} H_{f_i j}(t_{f_i}|\boldsymbol{X}_{f_i})}{k_j} \right)^{-k_j-d_{f_j}}}{1 - \prod_{j=1}^{J} \left(1 + \frac{H_{f_p j}(a_{f_p}|\boldsymbol{X}_{f_p})}{k_j} \right)^{-k_j}}.
$$

Maximum likelihood estimators of the parameters are obtained by maximizing the following ascertainment-corrected log-likelihood,

$$
\begin{aligned}
\ell_C(\boldsymbol{\theta}) &= \sum_{f=1}^{n} \log L_f(\boldsymbol{\theta}) - \sum_{f=1}^{n} \log A_f(\boldsymbol{\theta}) \\
&= \sum_{f=1}^{n} \left\{ \sum_{i=1}^{n_f} \sum_{j=1}^{J} \mathrm{I}(\delta_{f_i} = j) \log h_{f_i j}(t_{f_i}|\boldsymbol{X}_{f_i}) \right\} \\
&\quad + \sum_{f=1}^{n} \left\{ \sum_{j=1}^{J} \log((k_j + d_{f_j} - 1)!) - \log(k_j!) - (d_{f_j} - 1)\log(k_j) \right. \\
&\quad \left. - (k_j + d_{f_j}) \log\left(1 + \frac{\sum_{i=1}^{n_f} H_{f_i j}(t_{f_i}|\boldsymbol{X}_{f_i})}{k_j} \right) \right\} \\
&\quad - \sum_{f=1}^{n} \log\left(1 - \prod_{j=1}^{J} \left(1 + \frac{H_{f_p j}(a_{f_p}|\boldsymbol{X}_{f_p})}{k_j} \right)^{-k_j} \right).
\end{aligned}
\tag{3.9}
$$

3.3 Cause-specific penetrance function with time-varying covariates

Our main interest is to estimate the jth cause-specific cumulative incidence function $F_j(\cdot)$ or cause-specific penetrance. Conditional on the random frailty variable z_{f_j}, the cause-specific penetrance is expressed as:

$$\begin{aligned} F_{f_ij}(t|\boldsymbol{X}_{f_i}, z_{f_j}) &= P(T_{f_i} \leq t, \delta_{f_i} = j|\boldsymbol{X}_{f_i}, z_{f_j}) \\ &= \int_0^t h_{f_ij}(u|\boldsymbol{X}_{f_i}, z_{f_j}) S_{f_i}(u|\boldsymbol{X}_{f_i}, z_{f_j}) du \\ &= \int_0^t h_{f_ij}(u|\boldsymbol{X}_{f_i}, z_{f_j}) \exp\{-\sum_{j=1}^{J} H_{f_ij}(u|\boldsymbol{X}_{f_i}, z_{f_j})\} du. \end{aligned}$$

Since z_{f_j} is not observable, we obtain the marginal cause-specific penetrance function by integrating over the frailty distribution $g(z_{f_j})$. The marginal cause-specific cumulative incidence function for event j is obtained in the presence of J competing events for individual i in family f only given the vector of covariates $\boldsymbol{X}_{f_i}$:

$$F_{f_ij}(t|\boldsymbol{X}_{f_i}) = \int \cdots \int_J \int_0^t h_{f_ij}(u|\boldsymbol{X}_{f_i}, z_{f_j}) \cdot S_{f_i}(u|\boldsymbol{X}_{f_i}, z_{f_j}) \cdot g(z_{f_1}) \ldots g(z_{f_J}) \cdot du \cdot dz_{f_1} \ldots dz_{f_J}.$$

On the simplest case with only 2 competing events, $J = 2$, the cause-specific penetrance from cause 1, $j = 1$, omitting $\boldsymbol{X}_{f_i}$ for simplicity, can be expressed under $z_{f_j} \sim \text{Gamma}(k_j, \frac{1}{k_j})$:

$$
\begin{aligned}
F_{f_i1}(t|\boldsymbol{X}_{f_i}) &= \int_0^\infty \int_0^\infty \int_0^t h_{f_i1}(u|z_{f_1}) S_{f_i}(u|z_{f1}, z_{f_2}) g(z_{f_1}) g(z_{f_2}) du dz_{f_1} dz_{f_2} \\
&= \int_0^\infty \int_0^\infty \int_0^t h_{f_i1}(u) z_{f_1} \exp\Big\{ -H_{f_i1}(u) z_{f_1} - H_{f_i2}(u) z_{f_2} \Big\} g(z_{f_1}) g(z_{f_2}) du dz_{f_1} dz_{f_2} \\
&= \int_0^t h_{f_i1}(u) \iint z_{f_1} \exp\big\{ -H_{f_i1}(u) z_{f_1} - H_{f_i2}(u) z_{f_2} \big\} g(z_{f_1}) g(z_{f_2}) dz_{f_1} dz_{f_2} du \\
&= \int_0^t h_{f_i1}(u) \int \exp\Big\{ -H_{f_i2}(u) z_{f_2} \Big\} \int \exp\Big\{ -H_{f_i1}(u) z_{f_1} \Big\} g(z_{f_1}) dz_{f_1} g(z_{f_2}) dz_{f_2} du \\
&= \int_0^t h_{f_i1}(u) \cdot -\phi^{(1)}\Big(H_{f_i1}(u) \Big) \int \exp\Big\{ -H_{f_i2}(u) z_{f_2} \Big\} g(z_{f_2}) dz_{f_2} du \\
&= \int_0^t h_{f_i1}(u) \Big(1 + \frac{H_{f_i1}(u)}{k_1} \Big)^{-k_1-1} \Big(1 + \frac{H_{f_i2}(u)}{k_2} \Big)^{-k_2} du,
\end{aligned}
\tag{3.10}
$$

where we utilize Laplace transform $\phi(\cdot)$ of the frailty distribution $g(z_{f_j})$ and its dth derivative, $\phi(\cdot)^{(d)}$, similarly to the equations (3.7) and (3.8). Furthermore, we derive cause-specific penetrance function with TVC for the shared frailty competing risks model presented in (3.10). On the simplest case with only 2 competing events, $J = 2$, the cause-specific penetrance from cause 1, $j = 1$, for individual i in family f follows

$$
F_{f_i1}(t|X_{f_i}(t,t_x)) = \int_0^t h_{f_i1}(u|X_{f_i}(u,t_x)) \Big(1 + \frac{H_{f_i1}(u|X_{f_i}(u,t_x))}{k_1} \Big)^{-k_1-1} \Big(1 + \frac{H_{f_i2}(u|X_{f_i}(u,t_x))}{k_2} \Big)^{-k_2} du. \tag{3.11}
$$

If $t < t_x$, regardless of the structures of $X_{f_i}(t,t_x)$, equation (3.11) becomes

$$
F_{f_i1}(t|X_{f_i}(t)) = \int_0^t h_{01}(u) \Big(1 + \frac{H_{01}(u)}{k_1} \Big)^{-k_1-1} \Big(1 + \frac{H_{02}(u)}{k_2} \Big)^{-k_2} du.
$$

If $t \geq t_x$, equations (3.11) under PE, ED and CO follow respectively

$$F_{f_i1}(t|X_{f_i}(t,t_x)) = \int_0^t h_{01}(u)\exp\{\beta_1\} \times \left(1+\frac{H_{01}(t_x)+(H_{01}(u)-H_{01}(t_x))\exp\{\beta_1\}}{k_1}\right)^{-k_1-1} \times \left(1+\frac{H_{02}(t_x)+(H_{02}(u)-H_{02}(t_x))\exp\{\beta_2\}}{k_2}\right)^{-k_2} du, \qquad \text{(PE)}$$

$$F_{f_i1}(t|X_{f_i}(t,t_x)) = \int_0^t h_{01}(u)\exp\left\{\beta_1 e^{-\eta_1(u-t_x)}\right\} \times \left(1+\frac{H_{01}(t_x)+\int_{t_x}^u h_{01}(s)\exp\left\{\beta_1 e^{-\eta_1(s-t_x)}\right\}ds}{k_1}\right)^{-k_1-1} \times \left(1+\frac{H_{02}(t_x)+\int_{t_x}^u h_{02}(s)\exp\left\{\beta_2 e^{-\eta_2(s-t_x)}\right\}ds}{k_2}\right)^{-k_2} du, \qquad \text{(ED)}$$

$$F_{f_i1}(t|X_{f_i}(t,t_x)) = \int_0^t h_{01}(u)\exp\left\{\beta_1 e^{-\eta_1(u-t_x)}+\eta_{0_1}\right\} \times \left(1+\frac{H_{01}(t_x)+\int_{t_x}^u h_{01}(s)\exp\left\{\beta_1 e^{-\eta_1(s-t_x)}+\eta_{0_1}\right\}ds}{k_1}\right)^{-k_1-1} \times \left(1+\frac{H_{02}(t_x)+\int_{t_x}^u h_{02}(s)\exp\left\{\beta_2 e^{-\eta_2(s-t_x)}+\eta_{0_2}\right\}ds}{k_2}\right)^{-k_2} du, \qquad \text{(CO)}$$

where β_1 and β_2 are the cause-specific TVC effect coefficients, η_1 and η_2 are the cause-specficic TVC decay rate parameters and η_{01} and η_{02} are the cause-specific TVC decay convergence parameters for event 1 and 2, respectively.

3.4 Variance estimation

Let $\boldsymbol{\theta}$ be the vector of the parameters which consists of baseline parameters for specifying baseline hazard functions, regression coefficients and frailty parameters. The variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ is estimated using a robust sandwich variance estimator,

$$Var(\hat{\boldsymbol{\theta}}) = I_o(\boldsymbol{\theta})^{-1} J(\boldsymbol{\theta}) I_o(\boldsymbol{\theta})^{-1},$$

where $I_o(\boldsymbol{\theta})$ is the observed information matrix and $J(\boldsymbol{\theta})$ is the expected information matrix. They can be obtained by

$$\begin{aligned} I_o(\boldsymbol{\theta}) &= -\frac{\partial^2 \ell_C(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} \\ J(\boldsymbol{\theta}) &= U^T(\boldsymbol{\theta}) U(\boldsymbol{\theta}) \\ U(\boldsymbol{\theta}) &= \sum_f \frac{\partial \log L_f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \sum_f \frac{\partial \log A_f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \end{aligned}$$

where $\ell_C(\boldsymbol{\theta})$ is the ascertainment corrected log-likelihood of the data as presented in the equation (3.9), $\log L_f(\boldsymbol{\theta})$ is the log-likelihood contribution of the family f in equation (3.7) and $\log A_f(\boldsymbol{\theta})$ is the log-ascertainment correction term for the family f in equation (3.8). $\widehat{Var}(\hat{\boldsymbol{\theta}})$ is obtained by evaluating $I_o(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$ at the maximum-likelihood estimate $\hat{\boldsymbol{\theta}}$.

The robust variance estimator for the cause-specfic penetrance estimate, $F_j(t|\hat{\boldsymbol{\theta}})$, is obtained using Delta method:

$$Var(F_j(t|\hat{\boldsymbol{\theta}})) = D^T_{\boldsymbol{\theta}}(t) Var(\hat{\boldsymbol{\theta}}) D_{\boldsymbol{\theta}}(t),$$

where $Var(\hat{\boldsymbol{\theta}})$ is the robust variance estimator for $\boldsymbol{\theta}$ and $D_{\boldsymbol{\theta}}(t)$ is the vector of partial derivatives of $F_j(t|\hat{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\theta}$ evaluated at $\hat{\boldsymbol{\theta}}$.

Chapter 4

Simulation

In this chapter, we describe how we simulated family data based on the cause-specific shared frailty competing risk model with time varying covariates (TVC). We evaluate the performance of the proposed model under varying degree of familial dependence and three different structures of TVC in terms of model parameter and penetrance estimation. In Section 4.1, we summarize objectives of the simulation. In Section 4.2, we describe the data generation process. In Section 4.3 we discuss the simulation setting and evaluation criteria for the model performance. In Section 4.4, we summarize and discuss the simulation results.

4.1 Objectives of the simulation study

The simulation study has three main objectives.

1. To assess the bias and precision of the shared frailty competing risks model under three different structures (PE, ED, CO) of the time varying covariates.

2. To assess the bias and precision of each PE, ED and CO model under low, medium and high familial dependence.

3. To assess the bias and precision of each PE, ED and CO model under 1000, 500 and 250 families.

We designed the simulation study to assess the performance of the model described in Chapter 3 under different scenarios. First, each dataset is generated with two competing events with PE, ED and CO models, each with low, medium and high familial dependence and the numbers of families being 1000, 500, 250. Therefore, we evaluate a total of $3\times3\times3$=27 different scenarios. For each scenario, the bias and efficiency of the model parameter estimators and penetrance estimators are evaluated based on 500 simulations and results are summarized in Section 4.3.

4.2 Data generation

Data were simulated with code modified from the R package 'FamEvent' (Choi et al., 2017). Generation of the cause-specific competing risks survival data is based on the algorithm proposed by Beyersmann et al. (2009). Data generation and analyses were performed using R version 3.4.3. We choose randomly chosen, different starting seeds to generate data for each scenario, therefore simulations are completely independent (Burton et al., 2006).

We consider the shared frailty competing risk model with TVC for two competing risks. For the covariates, we include one TIC and one TVC.

1. G: Binary mutation status TIC. If the individual is a mutation carrier, G takes value of 1 otherwise 0. We assume cause specific hazards for both competing events are affected by this variable.

2. X(t,t_s): Screening status TVC as a function of time t and the screening time t_s: X(t,t_s) = I$(t \geq t_s)$. We assume only the cause-specific hazard for event 1 is affected by this variable.

Therefore, the cause-specfic hazards functions for event 1 and event 2 take the following forms:

$$\begin{aligned} h_1(t|X(t,t_s),G,z_1) &= h_{01}(t)\exp\{\beta_{g,1}G + g(X(t,t_s))\}z_1 \\ h_2(t|G,z_2) &= h_{02}(t)\exp\{\beta_{g,2}G\}z_2, \end{aligned} \tag{4.1}$$

where $h_{01}(t)$ and $h_{02}(t)$ are the Weibull baseline hazard functions, z_1 and z_2 are the cause-specfic shared frailties, $\beta_{G,1}$ and $\beta_{G,2}$ are the mutation status covariate coefficients for event 1 and 2, respectively, and $g(X(t,t_s))$ is the effect of the screening TVC, $X(t,t_s)$, which takes the following form depending on the model:

$$g(X(t,t_s)) = \begin{cases} 0 & \text{if } t < t_s \text{ (PE,ED,CO)} \\ \beta_s & \text{if } t \geq t_s \text{ (PE)} \\ \beta_s \cdot \exp\{-\eta(t-t_s)\} & \text{if } t \geq t_s \text{ (ED)} \\ \beta_s \cdot \exp\{-\eta(t-t_s)\} + \eta_0 & \text{if } t \geq t_s \text{ (CO)} . \end{cases}$$

The algorithm for generating families takes the following three steps based on model (4.1). Parameters specified in the data generation process, such as the number of siblings for each generation in family pedigree and the current age distribution of the probands and other family members result in the family structure similar to the real data in Chapter 5.

– Step 1: Family structure

(a) For each family, we generate a three-generation pedigree. We fix two members in the first generation while we generate 2 to 5 siblings in the second and 0 to 2 siblings in the third generations from a truncated negative binomial distribution.

(b) Generate the current age of the proband, a_{f_p} from normal distribution with mean age of 45 and SD of 10. Then we generate the current ages of other family members, $\{a_{f_2},...,a_{f_i}\}$ for individual i, $i = 2,\ldots,n_f$, from a normal distribution. The current ages of the first generation are generated with the mean age equal to $a_{f_p}+20$ with SD of 1.5 years. The current ages of the second generation are generated from mean age equivalent to a_{f_p} with SD of 1.5 years. Finally, for the third generation, their current ages are generated with the mean age subtracted by 20 years from the minimum age of their parents.

(c) To generate the screening TVC, we first generate the screening ages t_s for all members of the family from a normal distribution with mean age of 40 and variance of

2 years. If $t_{s,f_i} > a_{f_i}$ we assume this individual does not experience screening.

(d) Generate shared frailties $\boldsymbol{z}_f = \{z_{f_1}, z_{f_2}\}$ for family f for two competing events. We assume z_{f_1} and z_{f_2} are independent and marginally follow the gamma distribution with shape parameter k_1 and the scale parameter $1/k_1$ for event 1 and k_2 and $1/k_2$ for event 2 respectively.

(e) Generate the mutation status variable G_{f_p} for the proband assuming all the probands are the mutation carriers, based on a dominant model with prespecified *BRCA1* mutation allele frequency of 0.0021. Other family members' mutation statuses are generated conditioning on the proband's mutation status from a Bernoulli distribution with a probability of success equal to $P(G_{f_i} = 1|G_{f_p})$. This probability depends only on the relationship between the proband and the ith member of the family by Mendelian inheritance laws.

– Step 2: Event times and event types

(a) Generate t_{f_i} from the overall survival function: Generate w following a uniform on $[0,1]$ and solve for t_{f_i} from $P(T_{f_i} > t_{f_i}|G_{f_i}, t_{s,f_i}, \boldsymbol{z}_f) = w$.

(b) Given t_{f_i}, we decide the event type δ_{f_i} among two competing events using the rate of the cause-specific hazards at t_{f_i}. Compute $h_1(t_{f_i}|G_{f_i}, t_{s,f_i}, \boldsymbol{z}_f)$, $h_2(t_{f_i}|G_{f_i}, t_{s,f_i}, \boldsymbol{z}_f)$ and $p = \frac{h_1}{h_1+h_2}$. Run a Bernoulli experiment with the probability of success p. If success, then $\delta_{f_i} = 1$ otherwise $\delta_{f_i} = 2$. If $t_{f_i} > a_{f_i}$ we regard this individual as censored and $\delta_{f_i} = 0$. Follow-up duration is defined from age 16 to a_{f_i} if the individual is right censored, otherwise it is from age 16 to t_{f_i}.

– Step 3: Ascertaiment condition for the family

(a) After generating the event times and types of the family members, keep the family if it satisfies the condition $t_{f_p} < a_{f_p}$. This condition mimics the population based design of the family studies (Gong and Whittemore, 2003) where probands are affected before their study entry age, a_{f_p}.

(b) Remove men in the pedigree since the real data only consists of women. Mean

pedigree size of 5 leads to the total number of individuals about 2500 when 500 families are generated, which agrees with *BRCA1* data.

To generate the data, we specify the following parameters:

(a) baseline hazard function parameters: λ_1 and ρ_1 for event 1, λ_2 and ρ_2 for event 2

(b) parameters involved in TIC: $\beta_{g,1}$ and $\beta_{g,2}$ as genetic effects for each event

(c) parameters involved in TVC: β_s as a screening effect for event 1, η for ED and CO, additional η_0 for CO

(d) familial dependence paramter: k_1 and k_2 for each event

4.3 Simulation settings and evaluation criteria

Data was generated under the shared frailty competing risks model with three different TVCs. Values of the parameters are set using estimates obtained from fitting model (4.1) to the data described in Chapter 5. Their values and the corresponding penetrance values are summarized in Table 4.1. We present the true cause-specific penetrances by age 40, 50, 60 and 70 for Carrier (G=1)/Non-carrier (G=0) and Screened (S=1)/Non-screened (S=0) subgroups. We generated data with three familial dependences for event 1 ($k_1 = 7, 3.5, 1$) with n=1000, 500 and 250 families. Low familial dependence structure is generated with $k_1 = 7$ which corresponds to Kendall's $\tau = 0.07$. Medium dependence is generated using $k_1 = 3.5$ so $\tau = 0.13$ and high dependence with $k_1 = 1$ so $\tau = 0.33$. Kendall's τ closer to 1 indicates higher dependence between failure times. k_2 is fixed with the estimated value in Table 4.1. We evaluate a total of 9 (3 number of families $\times$ 3 familial dependence levels) scenarios for each TVC model and performed 500 simulations for each scenario.

Model performances are assessed through Bias, Empirical Standard Error (ESE), Average Standard Error (ASE) and Empirical Coverage Probability (ECP). Table 4.2 summarizes the four criteria for evaluating the performance of the simulation results.

Table 4.1: Parameter values used for generating family data for our simulation study and the corresponding penetrance values based on the true parameter values.

	Model parameters			Cause-specific penetrance given $t_s = 35$								
	TVC type				TVC type				TVC type		
	PE	ED	CO	Event 1	PE	ED	CO	Event 2	PE	ED	CO
λ_1	0.008	0.008	0.008	$F_1(40;S=0,G=0)$	0.021	0.024	0.023	$F_2(40;S=0,G=0)$	0.005	0.006	0.006
ρ_1	2.405	2.300	2.329	$F_1(40;S=1,G=0)$	0.028	0.042	0.031	$F_2(40;S=1,G=0)$	0.005	0.006	0.006
λ_2	0.007	0.007	0.007	$F_1(40;S=0,G=1)$	0.133	0.142	0.165	$F_2(40;S=0,G=1)$	0.014	0.019	0.027
ρ_2	3.080	2.932	2.906	$F_1(40;S=1,G=1)$	0.178	0.231	0.210	$F_2(40;S=1,G=1)$	0.014	0.018	0.026
β_s	0.668	1.872	3.401	$F_1(50;S=0,G=0)$	0.045	0.051	0.050	$F_2(50;S=0,G=0)$	0.013	0.016	0.016
β_{g1}	1.952	1.858	2.078	$F_1(50;S=1,G=0)$	0.075	0.073	0.061	$F_2(50;S=1,G=0)$	0.013	0.015	0.016
β_{g2}	1.194	1.224	1.566	$F_1(50;S=0,G=1)$	0.265	0.273	0.313	$F_2(50;S=0,G=1)$	0.035	0.044	0.062
k_1	3.225	3.435	3.536	$F_1(50;S=1,G=1)$	0.394	0.359	0.365	$F_2(50;S=1,G=1)$	0.032	0.040	0.058
k_2	2.900	3.240	3.529	$F_1(60;S=0,G=0)$	0.080	0.089	0.086	$F_2(60;S=0,G=0)$	0.027	0.031	0.032
η	-	0.278	3.530	$F_1(60;S=1,G=0)$	0.139	0.109	0.103	$F_2(60;S=1,G=0)$	0.026	0.031	0.032
η_0	-	-	0.160	$F_1(60;S=0,G=1)$	0.410	0.413	0.461	$F_2(60;S=0,G=1)$	0.065	0.079	0.105
				$F_1(60;S=1,G=1)$	0.586	0.477	0.514	$F_2(60;S=1,G=1)$	0.054	0.070	0.096
				$F_1(70;S=0,G=0)$	0.124	0.134	0.131	$F_2(70;S=0,G=0)$	0.047	0.054	0.055
				$F_1(70;S=1,G=0)$	0.215	0.153	0.154	$F_2(70;S=1,G=0)$	0.045	0.053	0.054
				$F_1(70;S=0,G=1)$	0.542	0.536	0.581	$F_2(70;S=0,G=1)$	0.099	0.116	0.148
				$F_1(70;S=1,G=1)$	0.721	0.582	0.629	$F_2(70;S=1,G=1)$	0.075	0.102	0.132

Bias is the measure of accuracy obtained as the discrepency between the true value of the parameter, β, and the average parameter estimate over the B simulations performed, $\bar{\hat{\beta}} = \sum_i^B \hat{\beta}_i/B$. Simulations are designed to mimic the results that could have been obtained from a single study and therefore the uncertainty in the estimate of interest between simulations, is assessed by the ESE, which is calculated as the standard deviation of the estimates of interest from all simulations. Acceptable level of bias is considered to be between $\frac{1}{2}SE(\hat{\beta})$ to $2SE(\hat{\beta})$ (Burton et al., 2006). ASE is the average of the robust standard errors obtained from each simulation. The ESE should be close to the ASE if the estimates are unbiased (Schafer and Graham, 2002).

Empirical Coverage Probability (ECP) is the proportion of times 95% confidence interval (CI) defined as $\hat{\beta}_i \pm Z_{0.975}\text{SE}(\hat{\beta}_i)$ include true value β, for $i = 1,,,B$, where B is the number of simulations and $\hat{\beta}_i$ is the estimate of β from simulation i. 95% of the CIs obtained from simulations have to contain the true parameter value, leading to 5% false positive rate where we expect 5% of the CIs do not contain true value. Any deviation from 5% indicates that we expect more or less type I error. For example, ECP higher than 95% results in a loss of statistical power with type I error lower than 5% but too many

type II error. In contrast, ECP lower than 95% is obtained when more than 5% of the CIs do not contain the true value. It is more critical since it relates to the increased type I error, we are rejecting true hypothesis too many times in case of testing the null effect. Acceptable range of coverage probability can be computed from the nominal coverage probability (p) and the number of independent simulations (B), as $p \pm 2\text{SE}(p)$ where $\text{SE}(p) = \sqrt{p(1-p)/B}$ (Tang et al., 2005). For our simulation, we have $\text{SE}(p) = 0.0097$ leading to the acceptable range of ECP between 93.05% and 96.95% for the nominal 95% coverage probability ($p = 0.95$).

Table 4.2: Summary of evaluation criteria.

Evaluation criteria	Formula
Bias Difference between the average estimate and the true value.	$\bar{\hat{\beta}}$-β
Empirical Standard Error (ESE) Standard deviation of the estimates of interest from all simulations.	$\sqrt{[1/(B-1)]\sum_{i=1}^{B}(\hat{\beta}_i - \bar{\hat{\beta}})^2}$
Average Standard Error (ASE) Average of robust SEs from each simulation i, $i = 1, ..., B$.	$\sum_{i=1}^{B} SE(\hat{\beta}_i)/B$
Empirical Coverage Probability (ECP)	Proportion of times 95% confidence interval defined as $\hat{\beta}_i \pm Z_{0.975}\text{SE}(\hat{\beta}_i)$ include true value β, for $i = 1, , B$.

B is the number of simulations, β is the true value of the parameter, $\hat{\beta}_i$ is the estimate of β from simulation i,
$\bar{\hat{\beta}}$ is average of the parameter estimates across all the simulations,
$\text{SE}(\hat{\beta}_i)$ is the estimated robust standard error from simulation i.

4.4 Simulation results and discussions

Simulation results of 500 families settings are summarized in Tables 4.3 to 4.5. Results of 1000 and 250 families are presented in the Tables A.1 to A.6 in the Appendix A. In Figures 4.1 to 4.6, we visualize the results for all the scenarios considered. Boxplots of the model parameters estimates are presented in Figures A.1 to A.6. We compare the bias in the model parameters and penetrance estimations between frailty model and independent model assuming data is generated with familial clustering in Table A.7 to A.10.

4.4.1 Cause-specific penetrances

Table 4.3 represents the result from PE TVC model including bias, ESE, ASE and ECP of the model parameters and cause-specific penetrance estimates for event 1 and 2. Table 4.4 and 4.5 summarize the results from ED and CO models, respectively. All the TVC models work well in terms of penetrance bias and precision for the event 1 and 2; bias was negligible ($< 1\%$) and ECPs were on average close to the 0.95 nominal level, ranged between 0.93 and 0.97 regardless of familial dependence levels. ASE and ESE mostly agreed in PE and ED models but ESE tends to be higher than ASE in the CO model. In addition, for the penetrance by age 70 for mutation carrier and screened individuals, ASEs were higher than ESEs indicating confidence intervals are slightly conservative leading to ECPs higher than nominal 95%. Figures 4.1 to 4.3 visualize accuracy and precision of the penetrance estimates for mutation carriers from n = 1000, 500 and 250 families based on 500 simulations for each TVC model. Solid red lines are the true penetrances for the the screened given that the screening occurred at age 35. Solid blue lines are those of the non-screened. Four small circles on each solid line indicate the mean penetrance estimates at age 40, 50, 60 and 70 respectively, while error bars are mean penetrance estimates $\pm$ 1.96ASE. Close alignments of solid lines and the circles indicate models produced negligible biases. We note that the true penetrance decreases with in-

creasing familial dependence ceteris paribus. There was approximately 10% difference in penetrance by age 70 for the screened carrier between medium and high familial dependence across all the TVC models. For the PE model, difference in the penetrances by age 70 between screened and non-screened groups was larger (19% in low, 18% in medium and 13% in high familial dependence) than ED and CO models (5% in low and medium, 3% in high familial dependence).

Generally, we observe accuracy of the penetrance estimates decreases with smaller number of families comparing the bias from Tables 4.3 to 4.6 to Tables A.1 to A.6, however bias is negligible regardless of the number of families. Figures 4.1 to 4.3 show that vertical lengths of the error bars increase as we reduce the number of families from 1000 to 250 indicating the precision of penetrance estimation also decreases with smaller number of families. Condifence intervals for the screened are slightly larger than the non-screened when evaluated at the ages 40, 50, 60 and 70. Cause-specific penetrance for event 2 is much smaller than the event 1 in our simulation settings (10% vs 65% by age 70).

In PE models from Tables 4.3, high familial dependence results in the smaller penetrance bias compared to the medium and low familial dependence (0.06% vs 0.17% at age 70 for event 1). In CO models from Table 4.6, however high familial dependence leads to the largest bias (0.13% vs 0.07% at age 70 for event 1). In PE model, ECPs for the event 2 are slightly lower in the low familial dependence (92% in low vs 93% in medium vs 93% in high for the lowest ECP acquired) but ED and CO models do not exhibit this pattern. In 500 and 250 families settings, penetrance biases are the highest in high familial dependence across all TVC models.

4.4.2 Model parameters

The model parameter estimates are presented in upper rows of Tables 4.3 to 4.5. ASEs and ESEs for all the parameter estimates agree each other and ECPs are mostly within ac-

ceptable range between 0.93 and 0.97 except for the frailty parameters (ranged between 0.86 and 0.97). ASEs tend to be larger than ESE in CO model. Lowest ECP occurred in PE model for k_2 and generally it has worse coverage than k_1. Again, bias is negligible for all the parameters except for the frailty parameters. Figures 4.4 and 4.5 visualize the results for the screening effect parameter, β_s and mutation gene effect for event 1, $\beta_{g,1}$. Results for the mutation gene effect for event 2, $\beta_{g,1}$, are presented in Table A.7. Figures compare the bias and the confidence intervals for PE, ED and CO models across all the number of families and familial dependence. All the TVC models work relatively well in terms of bias and precision for β_s with 1000 families regardless of familial dependence levels. However, CO model tends to overestimate β_s with 250 families when familial dependence is low or high. $\beta_{g,1}$ estimates are unbiased regardless the TVC models and the number of families but $\beta_{g,2}$ estimates were slightly overestimated in 250 families.

Precision and accuracy of frailty parameter estimates are relatively poorer compared to the other parameters in the model. Figure 4.6 summarizes the simulation results for frailty paramter k_1. ASE increases substantially with the lower familial dependence across all TVC models. This pattern is especially prominent in PE model.

We had some convergence issues under CO model; about 2% and 6% of the total simulations did not converge, respectively, with 500 families and 250 families. Each simulation is successfully completed only if `optim` function returned convergence code 0. Those with error code 10, implying degeneracy of the Nelder-Mead simplex, were discarded and additional data was generated to achieve 500 replications. PE and ED model had no convergence issues across all the sample sizes and dependence levels. CO model with 1000 families had no convergence issue but in 500 families setting, the simulations did not converge about 1.9% $(28/1500)$ of total simulations. In 250 families setting, those numbers increase to about 5.9% $(89/1500)$. Generally, CO model is susceptible to bias when the number of families are lower.

Table 4.3: (500 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Permanent Exposure TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations.

	Permanent Exposure Model														
	True	$k_1 = 7$, $\tau = 0.07$				True	$k_1 = 3.5$, $\tau = 0.13$				True	$k_1 = 1$, $\tau = 0.33$			
	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP
$log(\lambda_1)$	-4.83	-0.01	0.06	0.06	0.95	-4.83	0.00	0.06	0.06	0.95	-4.83	0.00	0.06	0.06	0.94
$log(\rho_1)$	0.88	0.00	0.03	0.03	0.94	0.88	0.00	0.03	0.03	0.93	0.88	0.00	0.03	0.03	0.96
$log(\lambda_2)$	-4.96	-0.01	0.09	0.10	0.95	-4.96	-0.02	0.10	0.10	0.94	-4.96	-0.01	0.09	0.10	0.96
$log(\rho_2)$	1.12	0.00	0.07	0.07	0.95	1.12	0.00	0.07	0.07	0.95	1.12	0.00	0.06	0.07	0.96
β_s	0.67	0.01	0.11	0.11	0.95	0.67	0.00	0.10	0.11	0.96	0.67	0.00	0.11	0.11	0.96
β_{g1}	1.95	0.01	0.12	0.12	0.95	1.95	0.01	0.12	0.12	0.96	1.95	0.00	0.12	0.11	0.94
β_{g2}	1.19	0.03	0.23	0.23	0.96	1.19	0.03	0.24	0.23	0.95	1.19	0.02	0.22	0.24	0.96
$log(k_1)$	1.95	0.24	1.08	0.85	0.92	1.25	0.13	0.69	0.48	0.95	0.00	0.02	0.25	0.25	0.95
$log(k_2)$	1.06	0.62	2.17	1.38	0.80	1.06	0.72	2.20	1.41	0.84	1.06	0.61	2.05	1.46	0.86
Cause-specific penetrance (%) given $t_{sc} = 35$															
Event 1															
$F_1(40;S=0,G=0)$	2.06	-0.01	0.26	0.26	0.93	2.05	0.00	0.26	0.27	0.94	2.04	0.01	0.29	0.28	0.95
$F_1(40;S=1,G=0)$	2.85	0.00	0.35	0.35	0.95	2.84	0.00	0.36	0.36	0.94	2.81	0.02	0.39	0.39	0.94
$F_1(40;S=0,G=1)$	13.43	-0.06	1.06	1.06	0.95	13.31	0.02	1.14	1.12	0.96	12.72	-0.02	1.30	1.30	0.95
$F_1(40;S=1,G=1)$	18.06	0.02	1.42	1.46	0.95	17.83	0.05	1.59	1.53	0.94	16.80	-0.01	1.71	1.75	0.95
$F_1(50;S=0,G=0)$	4.53	-0.04	0.53	0.52	0.94	4.52	0.00	0.52	0.54	0.95	4.45	0.03	0.59	0.58	0.94
$F_1(50;S=1,G=0)$	7.56	0.00	0.91	0.92	0.95	7.52	0.00	0.91	0.94	0.95	7.32	0.05	0.99	0.99	0.95
$F_1(50;S=0,G=1)$	27.07	-0.14	1.84	1.83	0.95	26.56	0.06	1.97	1.97	0.96	24.38	-0.03	2.28	2.30	0.94
$F_1(50;S=1,G=1)$	40.70	0.03	3.04	3.08	0.96	39.60	0.09	3.24	3.20	0.95	35.11	0.00	3.33	3.47	0.95
$F_1(60;S=0,G=0)$	8.07	-0.07	0.91	0.89	0.94	8.02	0.00	0.88	0.92	0.94	7.80	0.05	1.00	0.97	0.94
$F_1(60;S=1,G=0)$	14.05	0.00	1.63	1.64	0.94	13.91	0.01	1.61	1.67	0.95	13.26	0.09	1.72	1.73	0.96
$F_1(60;S=0,G=1)$	42.42	-0.24	2.62	2.62	0.95	41.23	0.10	2.79	2.81	0.95	36.41	-0.02	3.19	3.25	0.94
$F_1(60;S=1,G=1)$	61.28	-0.09	3.85	3.86	0.95	58.99	0.09	4.07	4.06	0.95	50.29	-0.01	4.27	4.45	0.95
$F_1(70;S=0,G=0)$	12.56	-0.10	1.38	1.36	0.95	12.45	0.01	1.33	1.40	0.94	11.93	0.07	1.48	1.45	0.94
$F_1(70;S=1,G=0)$	21.92	-0.01	2.45	2.45	0.94	21.58	0.02	2.37	2.48	0.95	20.09	0.13	2.49	2.50	0.96
$F_1(70;S=0,G=1)$	56.52	-0.33	3.20	3.18	0.94	54.51	0.12	3.39	3.42	0.94	46.80	-0.02	3.84	3.92	0.95
$F_1(70;S=1,G=1)$	75.63	-0.23	3.75	3.74	0.94	72.59	0.03	4.08	4.06	0.94	61.08	-0.04	4.61	4.79	0.94
Event 2															
$F_2(40;S=0,G=0)$	0.46	-0.01	0.10	0.11	0.92	0.46	-0.01	0.11	0.11	0.94	0.46	-0.01	0.10	0.11	0.93
$F_2(40;S=1,G=0)$	0.46	-0.01	0.10	0.10	0.92	0.46	-0.01	0.11	0.11	0.94	0.46	-0.01	0.10	0.10	0.93
$F_2(40;S=0,G=1)$	1.40	0.01	0.22	0.22	0.94	1.41	0.00	0.21	0.22	0.95	1.41	-0.01	0.23	0.22	0.94
$F_2(40;S=1,G=1)$	1.38	0.00	0.21	0.21	0.93	1.38	0.00	0.20	0.21	0.94	1.39	-0.01	0.22	0.22	0.94
$F_2(50;S=0,G=0)$	1.27	-0.02	0.24	0.25	0.93	1.27	-0.02	0.25	0.25	0.93	1.27	-0.02	0.23	0.25	0.94
$F_2(50;S=1,G=0)$	1.25	-0.02	0.24	0.24	0.93	1.25	-0.02	0.25	0.24	0.93	1.25	-0.02	0.23	0.24	0.94
$F_2(50;S=0,G=1)$	3.53	0.01	0.42	0.42	0.94	3.54	-0.01	0.41	0.42	0.94	3.58	-0.01	0.47	0.45	0.93
$F_2(50;S=1,G=1)$	3.21	0.00	0.39	0.38	0.93	3.23	-0.01	0.37	0.39	0.94	3.31	-0.01	0.42	0.40	0.94
$F_2(60;S=0,G=0)$	2.67	-0.05	0.47	0.48	0.94	2.67	-0.05	0.49	0.48	0.93	2.67	-0.03	0.45	0.49	0.95
$F_2(60;S=1,G=0)$	2.57	-0.05	0.45	0.46	0.94	2.57	-0.05	0.48	0.47	0.93	2.58	-0.03	0.44	0.48	0.95
$F_2(60;S=0,G=1)$	6.48	0.02	0.74	0.73	0.95	6.53	-0.02	0.73	0.74	0.94	6.74	0.00	0.82	0.79	0.94
$F_2(60;S=1,G=1)$	5.32	0.00	0.62	0.61	0.93	5.42	-0.02	0.61	0.62	0.94	5.83	-0.01	0.69	0.67	0.94
$F_2(70;S=0,G=0)$	4.73	-0.08	0.82	0.85	0.94	4.73	-0.08	0.87	0.85	0.93	4.74	-0.05	0.79	0.88	0.95
$F_2(70;S=1,G=0)$	4.45	-0.08	0.77	0.80	0.94	4.45	-0.08	0.82	0.80	0.93	4.49	-0.05	0.75	0.83	0.95
$F_2(70;S=0,G=1)$	9.68	0.04	1.16	1.15	0.94	9.85	-0.04	1.16	1.18	0.95	10.52	0.02	1.29	1.28	0.95
$F_2(70;S=1,G=1)$	7.12	0.01	0.91	0.89	0.94	7.42	-0.04	0.91	0.92	0.95	8.56	0.00	1.04	1.04	0.95

Table 4.4: (500 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Exponential Decay TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations.

Exponential Decay Model															
	True	$k_1 = 7$, $\tau = 0.07$				True	$k_1 = 3.5$, $\tau = 0.13$				True	$k_1 = 1$, $\tau = 0.33$			
	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP
$log(\lambda_1)$	-4.83	-0.01	0.05	0.06	0.96	-4.83	0.00	0.06	0.06	0.95	-4.83	0.00	0.06	0.06	0.96
$log(\rho_1)$	0.83	0.00	0.03	0.03	0.96	0.83	0.00	0.03	0.03	0.95	0.83	0.00	0.03	0.03	0.96
$log(\lambda_2)$	-4.96	0.00	0.09	0.09	0.95	-4.96	-0.01	0.09	0.09	0.96	-4.96	-0.01	0.09	0.09	0.95
$log(\rho_2)$	1.08	0.00	0.06	0.06	0.95	1.08	0.00	0.06	0.06	0.95	1.08	0.00	0.06	0.06	0.95
β_s	1.87	0.03	0.25	0.25	0.94	1.87	-0.01	0.25	0.25	0.95	1.87	0.03	0.24	0.24	0.94
β_{g1}	1.86	0.00	0.12	0.12	0.96	1.86	0.01	0.11	0.12	0.95	1.86	0.01	0.11	0.11	0.94
β_{g2}	1.22	0.01	0.20	0.21	0.95	1.22	0.03	0.22	0.21	0.96	1.22	0.02	0.21	0.22	0.96
$log(k_1)$	1.95	0.23	0.99	0.88	0.93	1.25	0.08	0.49	0.48	0.97	0.00	0.02	0.23	0.24	0.96
$log(k_2)$	1.18	0.51	2.04	1.18	0.85	1.18	0.53	1.70	1.26	0.84	1.18	0.48	1.47	1.28	0.84
$log(\eta)$	-1.28	0.02	0.32	0.31	0.94	-1.28	0.00	0.32	0.31	0.94	-1.28	0.03	0.30	0.30	0.94
Cause-specific penetrance (%) given $t_{sc} = 35$															
Event 1															
$F_1(40; S = 0, G = 0)$	2.43	0.00	0.29	0.30	0.95	2.42	0.00	0.30	0.30	0.95	2.40	-0.01	0.32	0.31	0.93
$F_1(40; S = 1, G = 0)$	4.21	0.01	0.59	0.59	0.95	4.20	-0.02	0.59	0.59	0.93	4.14	-0.02	0.57	0.60	0.95
$F_1(40; S = 0, G = 1)$	14.36	-0.05	1.01	1.10	0.96	14.21	0.04	1.14	1.16	0.95	13.55	0.02	1.29	1.32	0.97
$F_1(40; S = 1, G = 1)$	23.47	-0.03	2.16	2.28	0.95	23.09	-0.08	2.26	2.30	0.95	21.41	0.03	2.25	2.37	0.95
$F_1(50; S = 0, G = 0)$	5.16	-0.01	0.57	0.58	0.94	5.14	0.00	0.58	0.59	0.94	5.04	-0.02	0.62	0.62	0.94
$F_1(50; S = 1, G = 0)$	7.30	0.04	0.95	0.94	0.94	7.26	0.03	0.94	0.95	0.95	7.08	-0.01	0.90	0.96	0.96
$F_1(50; S = 0, G = 1)$	27.84	-0.13	1.66	1.84	0.97	27.30	0.03	1.87	1.96	0.96	25.01	0.05	2.19	2.25	0.95
$F_1(50; S = 1, G = 1)$	36.87	0.04	2.84	3.04	0.96	35.95	0.12	2.99	3.09	0.95	32.16	0.13	2.93	3.18	0.97
$F_1(60; S = 0, G = 0)$	8.91	-0.03	0.94	0.96	0.94	8.85	-0.01	0.95	0.97	0.94	8.59	-0.03	1.01	1.01	0.94
$F_1(60; S = 1, G = 0)$	10.98	0.05	1.26	1.26	0.94	10.89	0.05	1.24	1.27	0.94	10.49	0.00	1.23	1.28	0.97
$F_1(60; S = 0, G = 1)$	42.49	-0.21	2.26	2.56	0.97	41.29	0.00	2.55	2.73	0.97	36.44	0.10	3.00	3.09	0.95
$F_1(60; S = 1, G = 1)$	49.34	0.00	2.94	3.22	0.97	47.76	0.16	3.16	3.36	0.97	41.55	0.20	3.33	3.61	0.96
$F_1(70; S = 0, G = 0)$	13.55	-0.05	1.39	1.42	0.94	13.42	-0.02	1.41	1.44	0.94	12.82	-0.04	1.47	1.47	0.94
$F_1(70; S = 1, G = 0)$	15.49	0.03	1.64	1.64	0.94	15.32	0.05	1.62	1.66	0.94	14.54	0.00	1.61	1.68	0.97
$F_1(70; S = 0, G = 1)$	55.65	-0.28	2.70	3.07	0.97	53.68	-0.05	3.03	3.27	0.96	46.14	0.12	3.56	3.68	0.96
$F_1(70; S = 1, G = 1)$	60.49	-0.10	2.99	3.33	0.97	58.24	0.10	3.26	3.54	0.97	49.69	0.21	3.67	3.94	0.96
Event 2															
$F_2(40; S = 0, G = 0)$	0.59	0.00	0.12	0.12	0.93	0.59	-0.01	0.12	0.12	0.93	0.59	0.00	0.12	0.12	0.95
$F_2(40; S = 1, G = 0)$	0.59	0.00	0.12	0.12	0.93	0.59	-0.01	0.12	0.12	0.93	0.59	0.00	0.12	0.12	0.95
$F_2(40; S = 0, G = 1)$	1.86	0.00	0.26	0.26	0.95	1.86	0.00	0.27	0.26	0.94	1.86	0.01	0.26	0.26	0.97
$F_2(40; S = 1, G = 1)$	1.78	0.00	0.25	0.24	0.95	1.78	0.00	0.25	0.25	0.93	1.80	0.00	0.24	0.25	0.97
$F_2(50; S = 0, G = 0)$	1.55	0.00	0.27	0.28	0.94	1.55	-0.02	0.27	0.28	0.93	1.56	-0.01	0.26	0.28	0.96
$F_2(50; S = 1, G = 0)$	1.52	0.00	0.27	0.27	0.94	1.53	-0.02	0.27	0.27	0.94	1.53	-0.01	0.26	0.28	0.96
$F_2(50; S = 0, G = 1)$	4.42	0.01	0.47	0.48	0.95	4.44	0.01	0.51	0.49	0.94	4.49	0.00	0.50	0.51	0.96
$F_2(50; S = 1, G = 1)$	3.99	0.00	0.42	0.43	0.95	4.02	0.00	0.45	0.44	0.94	4.12	-0.01	0.44	0.46	0.96
$F_2(60; S = 0, G = 0)$	3.14	0.00	0.52	0.53	0.94	3.14	-0.05	0.51	0.53	0.94	3.14	-0.03	0.50	0.54	0.96
$F_2(60; S = 1, G = 0)$	3.07	0.00	0.51	0.52	0.94	3.07	-0.05	0.50	0.52	0.94	3.07	-0.03	0.48	0.53	0.96
$F_2(60; S = 0, G = 1)$	7.81	0.03	0.78	0.81	0.96	7.87	0.02	0.86	0.82	0.93	8.13	-0.02	0.86	0.87	0.96
$F_2(60; S = 1, G = 1)$	6.90	0.00	0.69	0.71	0.96	6.99	0.00	0.74	0.72	0.95	7.38	-0.03	0.75	0.77	0.96
$F_2(70; S = 0, G = 0)$	5.39	0.01	0.90	0.91	0.95	5.39	-0.07	0.86	0.92	0.95	5.41	-0.05	0.85	0.93	0.96
$F_2(70; S = 1, G = 0)$	5.26	0.01	0.87	0.89	0.95	5.26	-0.07	0.83	0.89	0.95	5.28	-0.06	0.83	0.91	0.95
$F_2(70; S = 0, G = 1)$	11.38	0.05	1.18	1.22	0.96	11.57	0.04	1.29	1.24	0.95	12.34	-0.05	1.34	1.35	0.95
$F_2(70; S = 1, G = 1)$	9.97	0.01	1.05	1.09	0.96	10.22	0.01	1.12	1.12	0.95	11.20	-0.07	1.20	1.22	0.95

Table 4.5: (500 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Cox and Oakes TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations.

	Cox and Oakes Model														
	True	$k_1 = 7$, $\tau = 0.07$				True	$k_1 = 3.5$, $\tau = 0.13$				True	$k_1 = 1$, $\tau = 0.33$			
	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP
$log(\lambda_1)$	-4.83	0.00	0.05	0.06	0.95	-4.83	0.00	0.05	0.06	0.94	-4.83	0.00	0.05	0.06	0.96
$log(\rho_1)$	0.83	0.00	0.03	0.03	0.94	0.83	0.00	0.03	0.03	0.96	0.83	0.00	0.03	0.03	0.97
$log(\lambda_2)$	-4.96	0.00	0.07	0.09	0.95	-4.96	0.00	0.07	0.09	0.97	-4.96	0.00	0.08	0.09	0.95
$log(\rho_2)$	1.07	0.00	0.05	0.06	0.96	1.07	0.00	0.05	0.06	0.97	1.07	0.00	0.05	0.06	0.96
β_s	1.52	0.04	0.32	0.42	0.96	1.52	0.04	0.33	0.42	0.94	1.52	0.02	0.32	0.42	0.96
β_{g1}	2.08	0.01	0.10	0.12	0.94	2.08	0.01	0.10	0.12	0.95	2.08	0.01	0.09	0.11	0.96
β_{g2}	1.57	0.00	0.17	0.21	0.98	1.57	0.00	0.17	0.21	0.94	1.57	0.01	0.16	0.21	0.97
$log(k_1)$	1.95	0.20	0.74	0.86	0.91	1.25	0.10	0.39	0.46	0.96	0.00	0.02	0.18	0.22	0.97
$log(k_2)$	1.26	0.38	1.15	1.39	0.86	1.26	0.35	0.98	1.40	0.90	1.26	0.36	1.10	1.32	0.87
$log(\eta)$	-0.18	-0.02	0.50	0.58	0.90	-0.18	0.01	0.50	0.60	0.91	-0.18	-0.03	0.48	0.62	0.91
η_0	0.21	-0.02	0.12	0.14	0.95	0.21	-0.01	0.12	0.14	0.96	0.21	-0.02	0.12	0.14	0.95
Cause-specific penetrance (%) given $t_{sc} = 35$															
Event 1															
$F_1(40;S=0,G=0)$	2.43	0.01	0.28	0.29	0.95	2.42	0.00	0.29	0.30	0.94	2.40	0.00	0.28	0.30	0.95
$F_1(40;S=1,G=0)$	3.25	0.05	0.44	0.45	0.95	3.24	0.03	0.44	0.45	0.94	3.21	0.03	0.43	0.46	0.96
$F_1(40;S=0,G=1)$	17.45	0.09	1.24	1.21	0.95	17.23	0.09	1.25	1.26	0.94	16.27	0.08	1.40	1.43	0.95
$F_1(40;S=1,G=1)$	22.47	0.32	2.08	2.03	0.93	22.12	0.21	1.95	2.05	0.97	20.56	0.23	2.01	2.13	0.96
$F_1(50;S=0,G=0)$	5.15	0.01	0.54	0.58	0.95	5.13	0.01	0.58	0.58	0.94	5.04	0.02	0.56	0.60	0.96
$F_1(50;S=1,G=0)$	6.57	0.07	0.82	0.83	0.95	6.54	0.04	0.83	0.83	0.94	6.39	0.04	0.80	0.84	0.95
$F_1(50;S=0,G=1)$	32.90	0.12	1.97	1.95	0.96	32.16	0.16	2.09	2.05	0.94	29.06	0.19	2.27	2.34	0.96
$F_1(50;S=1,G=1)$	39.48	0.33	3.15	2.92	0.93	38.43	0.29	2.89	2.97	0.96	34.14	0.26	2.95	3.06	0.95
$F_1(60;S=0,G=0)$	8.91	0.02	0.90	0.96	0.96	8.85	0.01	0.97	0.96	0.95	8.58	0.04	0.92	0.98	0.96
$F_1(60;S=1,G=0)$	11.09	0.03	1.36	1.37	0.95	11.00	0.01	1.37	1.37	0.94	10.59	0.01	1.32	1.37	0.95
$F_1(60;S=0,G=1)$	48.48	0.11	2.59	2.59	0.95	46.94	0.23	2.79	2.73	0.94	40.89	0.28	2.94	3.07	0.96
$F_1(60;S=1,G=1)$	55.53	0.07	3.86	3.61	0.93	53.57	0.17	3.63	3.69	0.95	46.03	0.16	3.66	3.78	0.96
$F_1(70;S=0,G=0)$	13.54	0.02	1.36	1.42	0.95	13.41	0.02	1.44	1.43	0.95	12.81	0.08	1.34	1.43	0.96
$F_1(70;S=1,G=0)$	16.60	-0.04	2.00	2.03	0.95	16.41	-0.04	2.02	2.03	0.94	15.52	-0.02	1.89	1.99	0.95
$F_1(70;S=0,G=1)$	61.12	0.07	2.90	2.93	0.96	58.82	0.25	3.15	3.10	0.94	50.11	0.32	3.32	3.49	0.97
$F_1(70;S=1,G=1)$	67.55	-0.15	3.94	3.73	0.93	64.90	0.06	3.80	3.86	0.95	54.88	0.09	3.94	4.09	0.95
Event 2															
$F_2(40;S=0,G=0)$	0.62	0.01	0.12	0.13	0.96	0.62	0.01	0.12	0.13	0.95	0.62	-0.01	0.11	0.13	0.97
$F_2(40;S=1,G=0)$	0.62	0.01	0.12	0.13	0.96	0.62	0.01	0.12	0.13	0.95	0.62	-0.01	0.11	0.13	0.97
$F_2(40;S=0,G=1)$	2.66	0.00	0.31	0.33	0.96	2.67	0.00	0.32	0.33	0.96	2.68	-0.01	0.33	0.34	0.94
$F_2(40;S=1,G=1)$	2.58	0.00	0.30	0.32	0.96	2.59	0.00	0.31	0.32	0.96	2.61	-0.01	0.32	0.33	0.95
$F_2(50;S=0,G=0)$	1.61	0.01	0.27	0.29	0.95	1.61	0.01	0.27	0.29	0.95	1.61	-0.01	0.26	0.29	0.95
$F_2(50;S=1,G=0)$	1.59	0.01	0.27	0.28	0.95	1.59	0.01	0.27	0.28	0.95	1.59	-0.01	0.26	0.28	0.95
$F_2(50;S=0,G=1)$	6.11	-0.02	0.56	0.59	0.96	6.13	0.00	0.60	0.60	0.96	6.24	-0.03	0.62	0.63	0.95
$F_2(50;S=1,G=1)$	5.66	-0.03	0.52	0.54	0.96	5.70	-0.01	0.55	0.55	0.96	5.86	-0.04	0.58	0.58	0.95
$F_2(60;S=0,G=0)$	3.23	0.02	0.51	0.54	0.94	3.23	0.02	0.52	0.55	0.95	3.24	-0.02	0.51	0.55	0.95
$F_2(60;S=1,G=0)$	3.18	0.02	0.50	0.53	0.95	3.18	0.02	0.51	0.54	0.96	3.18	-0.02	0.50	0.54	0.96
$F_2(60;S=0,G=1)$	10.30	-0.03	0.88	0.95	0.97	10.42	0.00	0.97	0.96	0.95	10.90	-0.05	1.02	1.03	0.95
$F_2(60;S=1,G=1)$	9.22	-0.06	0.83	0.87	0.94	9.37	-0.02	0.89	0.89	0.96	10.00	-0.06	0.93	0.94	0.95
$F_2(70;S=0,G=0)$	5.53	0.04	0.87	0.93	0.95	5.53	0.05	0.88	0.93	0.95	5.55	-0.02	0.89	0.95	0.95
$F_2(70;S=1,G=0)$	5.39	0.03	0.85	0.90	0.95	5.39	0.04	0.85	0.91	0.95	5.42	-0.02	0.87	0.93	0.95
$F_2(70;S=0,G=1)$	14.27	-0.06	1.24	1.37	0.98	14.61	-0.02	1.38	1.41	0.94	15.91	-0.08	1.51	1.55	0.95
$F_2(70;S=1,G=1)$	12.35	-0.06	1.22	1.28	0.96	12.77	-0.02	1.31	1.32	0.95	14.36	-0.07	1.41	1.45	0.95

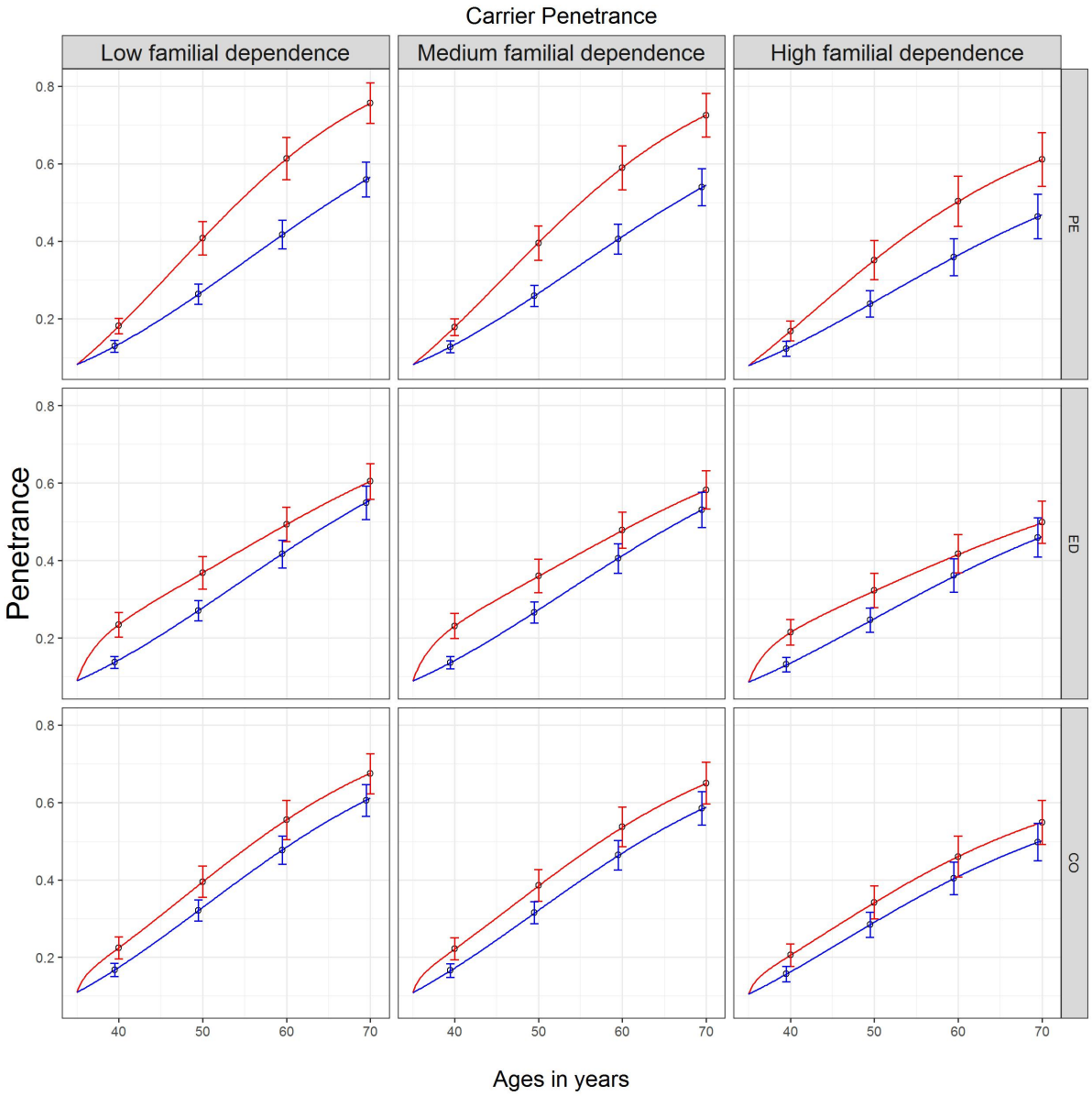

Figure 4.1: Accuracy and precision of the penetrance estimates for mutation carriers from n = 1000 families based on 500 simulations. The red line represents the true penetrance of the screened group and the blue line represents non-screened group.

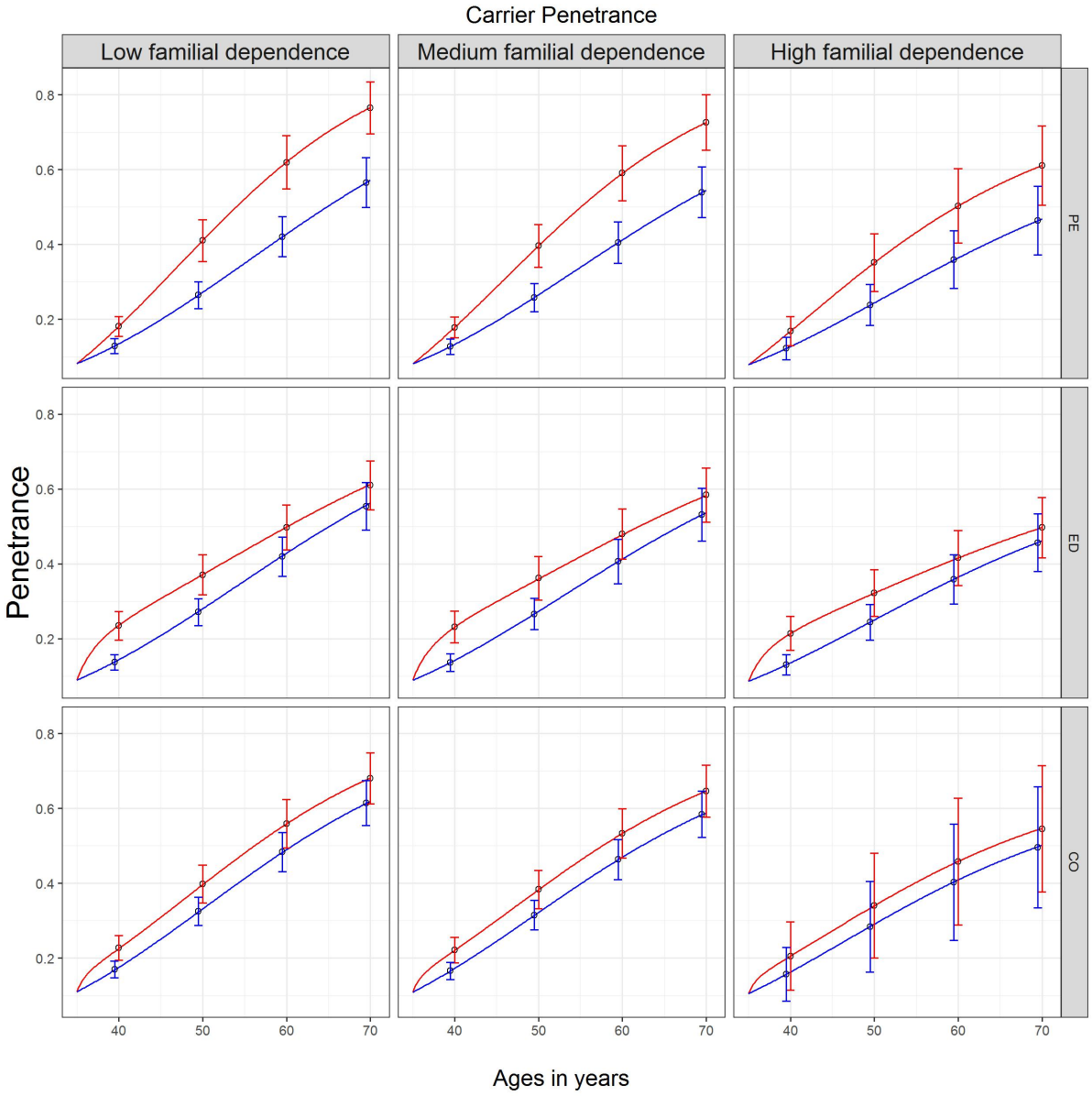

Figure 4.2: Accuracy and precision of the penetrance estimates for mutation carriers from n = 500 families based on 500 simulations. The red line represents the true penetrance of the screened group and the blue line represents non-screened group.

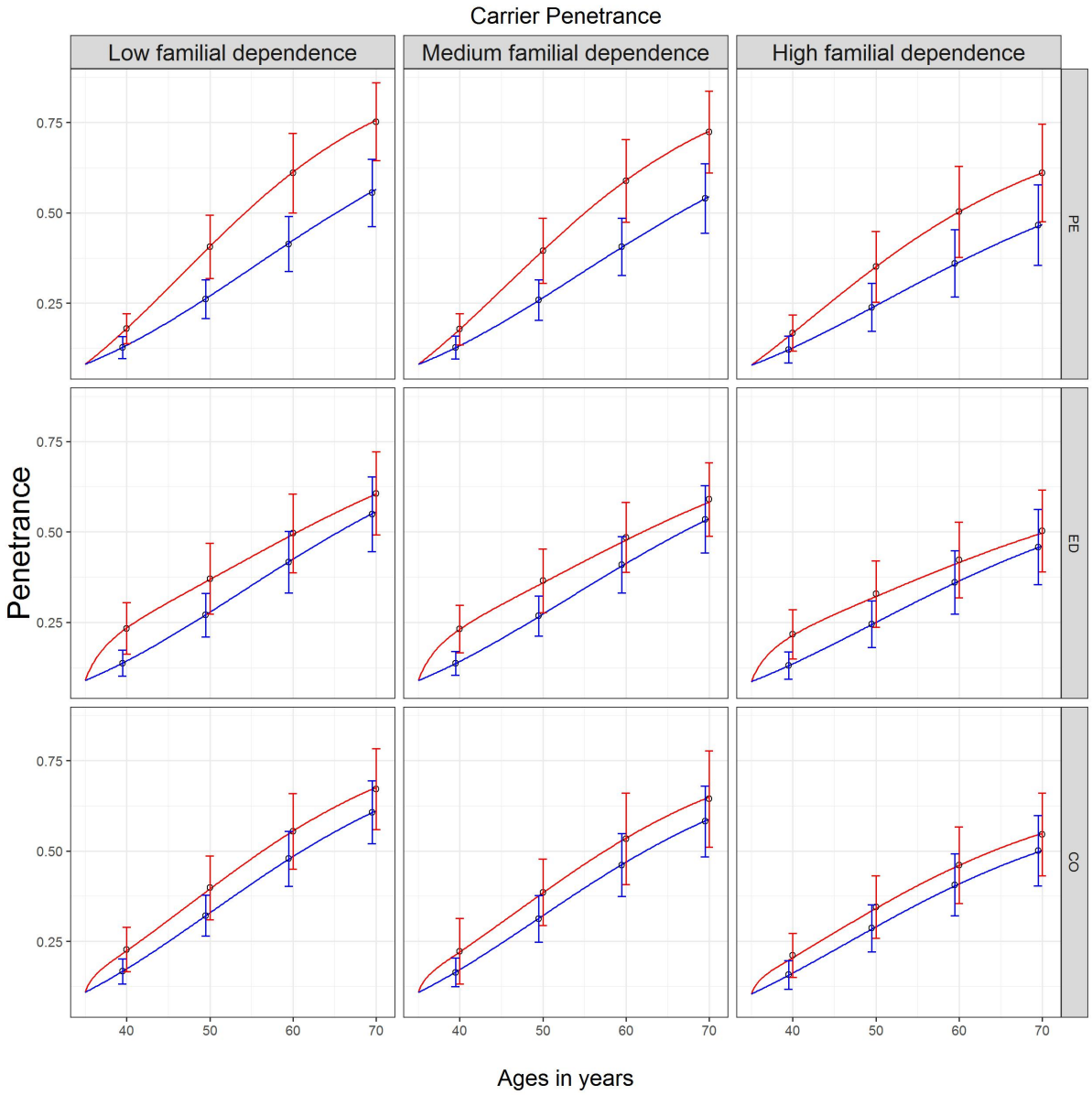

Figure 4.3: Accuracy and precision of the penetrance estimates for mutation carriers from n = 250 families based on 500 simulations. The red line represents the true penetrance of the screened group and the blue line represents non-screened group.

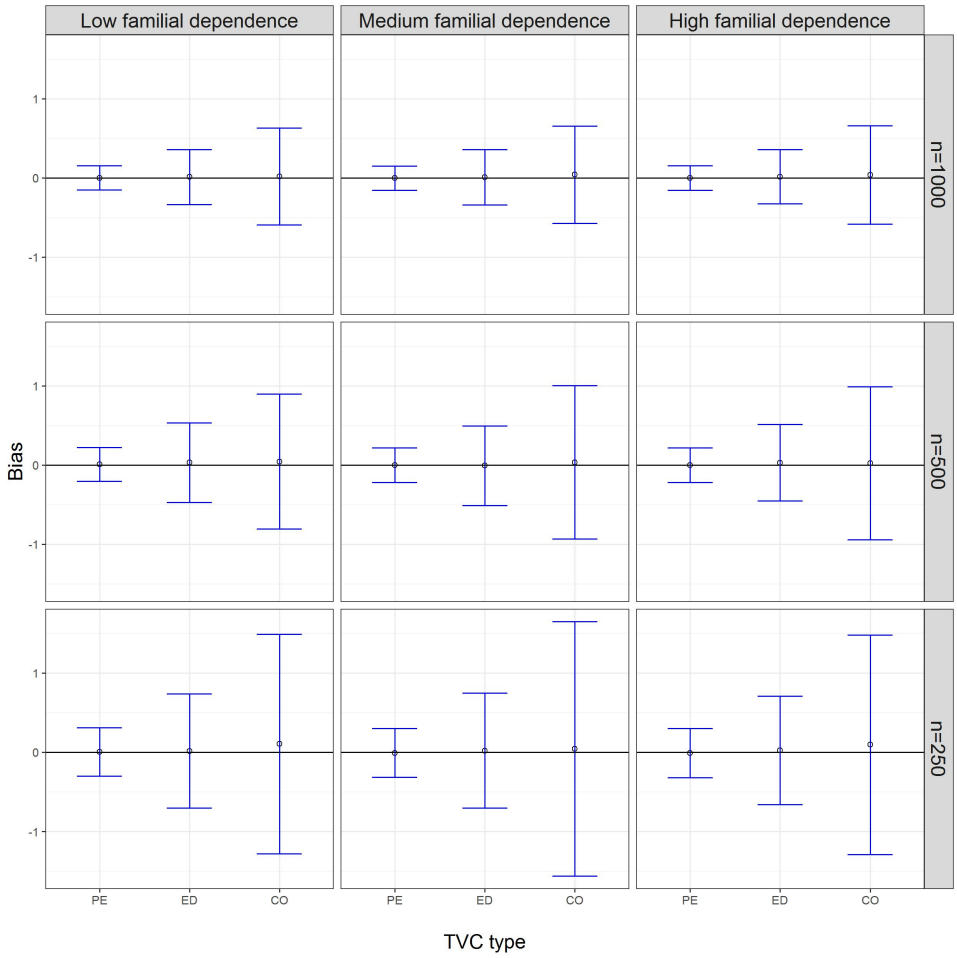

Figure 4.4: Bias and precision of the parameter estimates for screen effect, β_s, expressed as bias $\pm$ 1.96ASE, based on 500 simulations. True value of the β_s is 0.668 for PE, 1.872 for ED and 3.401 for CO.

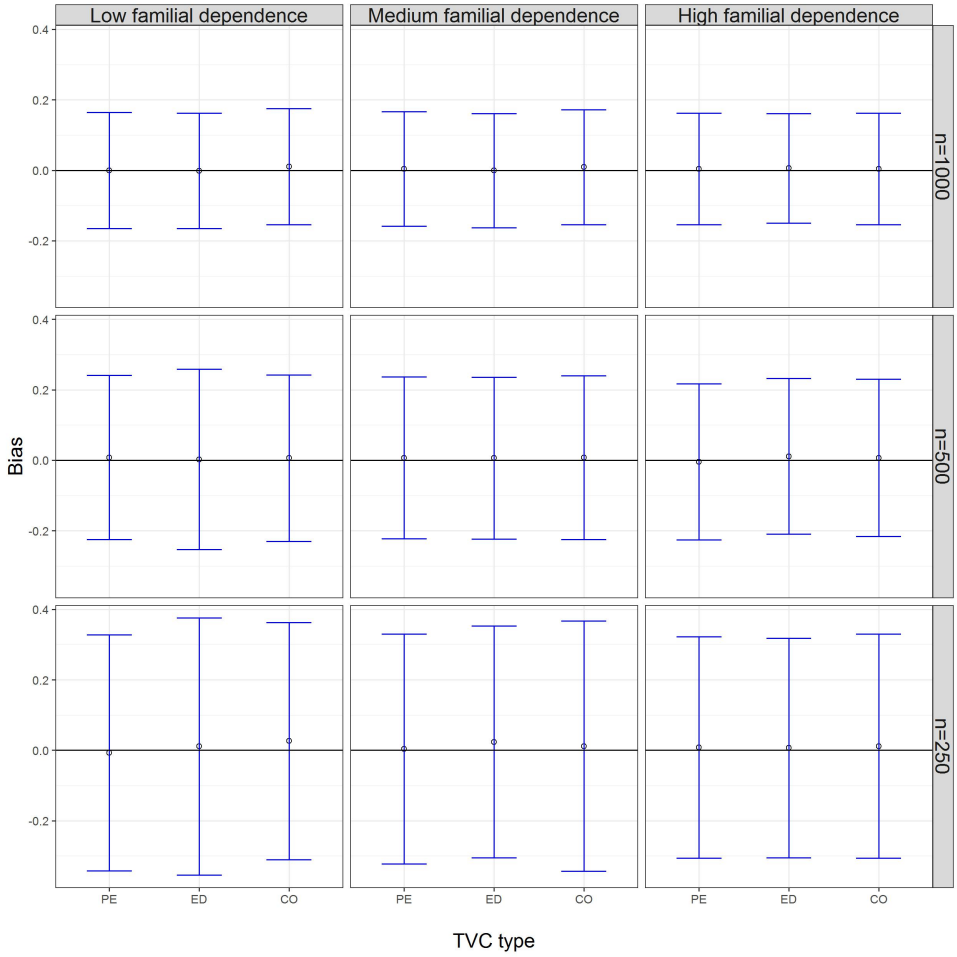

Figure 4.5: Bias and precision of the parameter estimates for mutation effect for event 1, $\beta_{g,1}$, expressed as bias $\pm$ 1.96ASE, based on 500 simulations. True value of the $\beta_{g,1}$ is 1.952 for PE, 1.858 for ED and 2.078 for CO.

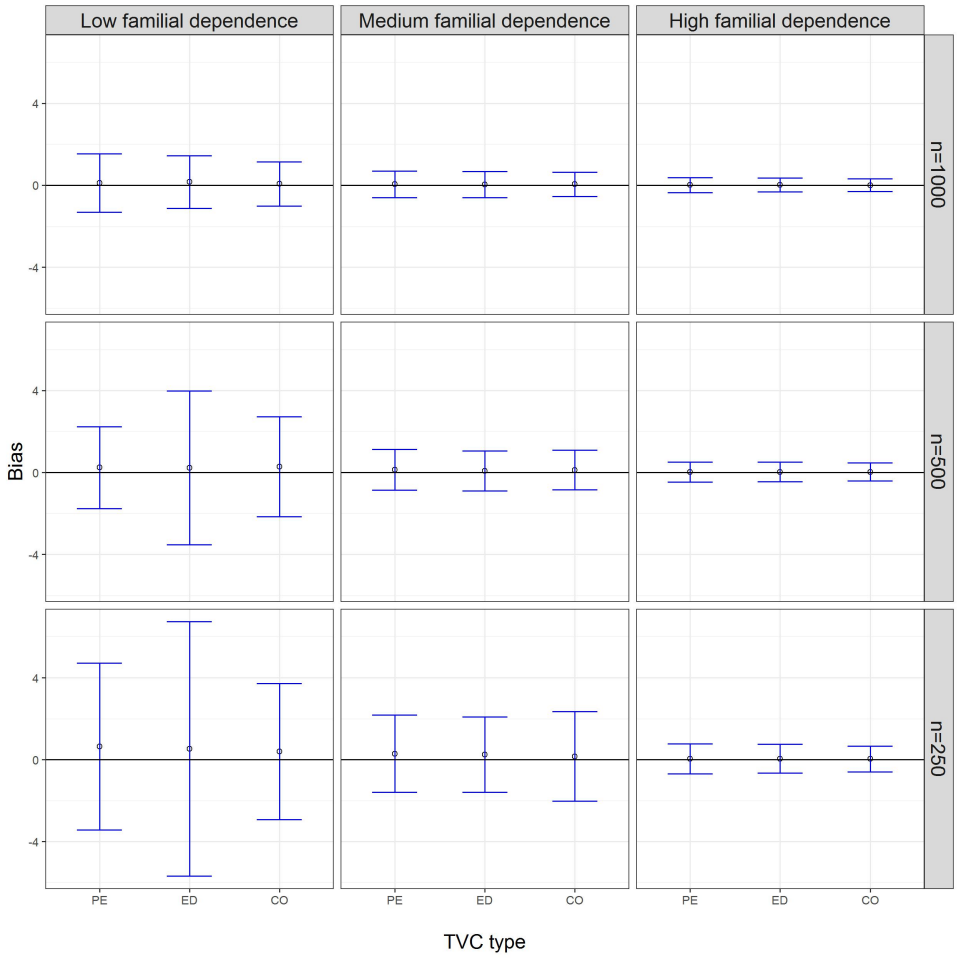

Figure 4.6: Bias and precision of the parameter estimates for frailty parameter for event 1, k_1, expressed as bias $\pm$ 1.96ASE, based on 500 simulations. True value of the $\log(k_1)$ is 2.30 for low, 1.25 for medium and 0 for high familial dependence.

Chapter 5

Application to hereditary breast and ovarian cancer families

This chapter presents the analysis of the *BRCA*1/2 mutation positive family data. Section 5.1 introduces the data and descriptive statistics. Then in Section 5.2, we describe the specification of the model fitted. Fianlly in Section 5.3, we summarize and discuss the modelling results. Our main focus is on the relative risk estimation of woman who undergoes Bilateral Oophorectomy (BO) and Mammographic Screenings (MS) and the cause-specific penetrance estimation of the event of interest which is breast cancer, accounting for the competing risks, familial dependence and the ascertainment bias. Comparison of the models with different TVC functions is also presented. We used custom codes based on *FamEvent* R package to fit the model. Some of the key R functions are listed in Appendix B.

5.1 The data

Our study data consists of 876 *BRCA* positive families recruited through the Breast Cancer Family Registries (BCFRs). The BCFRs database (John et al., 2004) was established

in 1995 with six participating sites from USA, Austrailia and Canada including Ontario Cancer Care. BCFRs enrolled most of the families from 1996 to 2000 while continuing to recruit additional families satisfying its criteria. Families are registered if *BRCA1* or *BRCA2* mutations pass down to further generations (segregating mutation), exhibit multiple cases of breast or ovarian cancer, have Ashkenazi Jewish ancestry or come from specific racial and ethnic groups. For the population based families, each family includes the proband, i.e. the initial member of the family to come under study, as well as the first and the second degree relatives. Data have extensive information on the family members including the ages at breast/ovarian cancer diagnosis, study entry, screenings including Magnetic Resonance Imaging (MRI) and mammography and surgeries (Mastectomy and Oophorectomy) and mutation status in *BRCA1/2* gene. The data consist of 4575 women including 498 *BRCA1* families (2650 women) and 378 *BRCA2* families (1925 women). Among 4575 women, 1639 women develop breast cancer (BC) as the first event. Breast cancer is the event of interest and the ovarian cancer (OC) and the death are considered as the competing events. Women are right censored if they do not experience any of the endpoints during follow-up. Figure 5.1 summarizes the data we have for analysis.

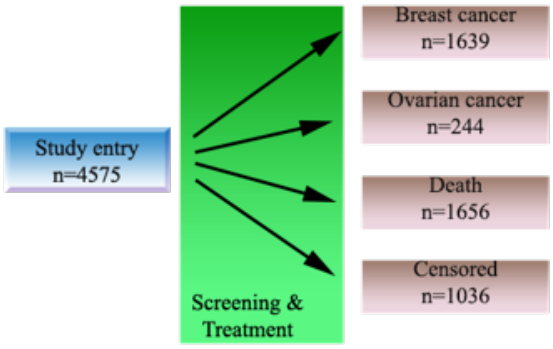

Figure 5.1: Total number of women entered the study with competing risks, *BRCA1/2* combined.

We separately analyzed the *BRCA1* mutation positive families and the *BRCA2* mutation positive families. Detailed descriptive statistics of the *BRCA1* and the *BRCA2* families are presented in Tables 5.1 and 5.2. We only considered those screenings and surgeries which occured before any events (cancers or death). The proportion of the women who have OC as the first cancer is much lower than that of BC in both *BRCA1/2*. (34.9% vs

6.9% in *BRCA1*, 37.1% vs 3.2% in *BRCA2*). We include two surgeries into the analysis: Bilateral Mastectomy (BM) and Bilateral Oophorectomy (BO). If the woman has BM, She cannot develop breast cancer after the surgery since breast tissue is missing while BO similarly prevents ovarian cancer. This is reflected in the data as the breast cancer cohort has no BM and ovarian cancer cohort has no BO in Tables 5.1 and 5.2. Also, we still have very limited number of women who undergo BO among the breast cancer cohort (4.3% in *BRCA1*, 4.5% in *BRCA2*) and BM among the ovarian cancer cohort (0.6% in *BRCA1*, 0% in *BRCA2*). The former case is of more interest concerning the evidence of the negative association between the oophorectomy and the etiology of breast cancer (Eisen et al., 2005).

Table 5.1: Descriptive statistics of *BRCA1* positive families.

	Breast Cancer	Ovarian Cancer	Death	Censored	Total
N (%)	924 (34.9%)	182 (6.9%)	958 (36.2%)	586 (22.1%)	2650
Event age					
mean (SD)	44.2 (12.0)	53.0 (11.5)	70.5 (17.9)	50.9 (16.2)	55.8 (19.1)
min, max	21.0, 86.0	28.0, 89.0	18.5, 102.5	18.1, 95.0	18.1, 102.5
BRCA Mutation					
Noncarrier	29 (3.1%)	4 (2.2%)	14 (1.5%)	229 (39.1%)	276 (10.4%)
Carrier	483 (52.3%)	55 (30.2%)	16 (1.7%)	192 (32.8%)	746 (28.2%)
Untested	412 (44.6%)	123 (67.6%)	928 (96.9%)	165 (28.2%)	1628 (61.4%)
# of screening					
0	722 (78.1%)	158 (86.8%)	944 (98.5%)	257 (43.9%)	2081 (78.5%)
1	157 (17.0%)	19 (10.4%)	7 (0.7%)	174 (29.7%)	357 (13.5%)
2	30 (3.3%)	4 (2.2%)	3 (0.3%)	58 (10.0%)	95 (3.6%)
3+	15 (1.6%)	1 (0.6%)	4 (0.4%)	97 (16.6%)	117 (4.4%)
Surgery					
None	884 (95.7%)	181 (99.5%)	946 (98.8%)	441 (75.3%)	2452 (92.5%)
BM	0 (0.0%)	1 (0.6%)	3 (0.3%)	16 (2.7%)	20 (0.8%)
BO	40 (4.3%)	0 (0.0%)	9 (0.9%)	85 (14.5%)	134 (5.1%)
Both	0 (0.0%)	0 (0.0%)	0 (0.0%)	44 (7.5%)	44 (1.7%)

BO for Bilateral Oophorectomy, BM for Bilateral Mastectomy and SD for Standard Deviation.

Most women were not screened (78.5% in *BRCA1*, 76.8% in *BRCA2*). If they were screened, they commonly have just one screening (13.5% in *BRCA1*, 13.7% in *BRCA2*).

A small proportion of women had multiple screens (8% in *BRCA1*, 9.5% in *BRCA2*) and most women who have multiple screenings have not yet developed any cancer or died.

Table 5.2: Descriptive statistics of *BRCA2* mutation positive analysis cohorts.

	Breast Cancer	Ovarian Cancer	Death	Censored	Total
N (%)	715 (37.1%)	62 (3.2%)	698 (36.3%)	450 (23.4%)	1925
Event age					
mean (SD)	47.9 (12.8)	54.3 (11.3)	69.9 (18.2)	51.8 (16.3)	57.0 (18.6)
min, max	21.8, 94.0	30.0, 80.0	16.5, 109.5	19.7, 97.0	16.5, 109.5
BRCA Mutation					
Noncarrier	31 (4.3%)	3 (4.8%)	17 (2.4%)	196 (43.6%)	247 (12.8%)
Carrier	373 (52.2%)	18 (29.0%)	10 (1.4%)	175 (38.9%)	576 (29.9%)
Untested	311 (43.5%)	41 (66.1%)	671 (96.1%)	79 (17.6%)	1102 (57.3%)
# of screening					
0	566 (79.2%)	58 (46.8%)	685 (98.1%)	170 (37.8%)	1479 (76.8%)
1	107 (15.0%)	4 (3.2%)	7 (1%)	145 (32.2%)	263 (13.7%)
2	26 (3.6%)	58 (46.8%)	4 (0.6%)	45 (10.0%)	75 (3.9%)
3+	16 (2.2%)	4 (3.2%)	2 (0.3%)	90 (20.0%)	108 (5.6%)
Surgery					
None	683 (95.5%)	62 (100.0%)	689 (98.7%)	335 (74.4%)	1769 (91.9%)
BM	0 (0.0%)	0 (0.0%)	0 (0.0%)	8 (1.8%)	8 (0.4%)
BO	32 (4.5%)	0 (0.0%)	9 (1.3%)	83 (18.4%)	124 (6.4%)
Both	0 (0.0%)	0 (0.0%)	0 (0.0%)	24 (5.3%)	24 (1.3%)

BO for Bilateral Oophorectomy, BM for Bilateral Mastectomy and SD for Standard Deviation.

Tables 5.3 and 5.4 present descriptive statistics for the probands of *BRCA1* and the *BRCA2* families. In terms of their study entry ages, event ages, *BRCA* mutation status and the number of screenings and surgeries. We classify the probands into three subgroups: BC, OC and Unaffected. BC subgroup consists of the probands who are affected by breast cancer before the study entry (77.5% in *BRCA1*, 79.6% in *BRCA2*). Their mean event age and study entry age are 43 years and 48 years respectively. Probands in OC subgroup are affected by ovarian cancer before study entry and their proportion is relatively small compared to BC subgroup (6.2% in *BRCA1*, 2.7% in *BRCA2*). Unaffected probands are included in the data without being affected by either cancer before the study entry. Probands are all mutation carriers in both *BRCA1* and *BRCA1* families.

The unaffected have more surgeries than BC or OC subgroups, in *BRCA1*, there are 2 BM, 23 BO and 21 cases with both of the surgeries. *BRCA2* unaffected probands have 1, 22 and 12 cases for BM, BO and both surgeries, respectively. We have very limited number of women who undergo BO among BC subgroup (4.7% in *BRCA1*, 4.0% in *BRCA2*). There is no BM among OC subgroup in both *BRCA1* and *BRCA2*. About a half of the probands do not have screenings (57.8% in *BRCA1*, 57.1% in *BRCA2*). They commonly have only one screening (28.9% in *BRCA1*, 29.6% in *BRCA2*). A small proportion of the probands has multiple screenings (13.3% in *BRCA1*, 13.2% in *BRCA2*).

In Table 5.5, we summarize the distribution of the ages at the first screening, BO and their gap times in *BRCA1* and *BRCA2* families. Gap times between consecutive screenings are also presented. Mean age of the first screenings is about 40 years in *BRCA1* and 43 years in *BRCA2*. Mean age of BO is about 46 years in *BRCA1* and 47 years in *BRCA2*. Mean gap time between the first and second screenings is about 9 years and that between the second and third is about 6 years.

Table 5.3: Descriptive statistics for the probands of *BRCA1* positive families.

	Breast Cancer	Ovarian Cancer	Unaffected	Total
N (%)	386 (77.5%)	31 (6.2%)	81 (16.3%)	498
Study entry age				
mean (SD)	44.7 (9.9)	53.6 (9.9)	39.9 (10.1)	44.5 (10.3)
min, max	24.0, 81.5	39.6, 79.8	22.4, 71.0	22.4, 81.5
Event age				
mean (SD)	40.1 (8.2)	48.3 (8.4)	48.4 (11.5)	41.9 (9.5)
min, max	21.9, 66.7	36.0, 79.0	22.4, 79.5	21.9, 79.5
BRCA Mutation				
Carrier	386 (100.0%)	31 (100.0%)	81 (100.0%)	498 (100.0%)
Noncarrier	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Untested	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
# of screening				
0	249 (64.5%)	20 (64.5%)	19 (23.5%)	288 (57.8%)
1	114 (29.5%)	9 (29.0%)	21 (25.9%)	144 (28.9%)
2	17 (4.4%)	2 (6.5%)	12 (14.8%)	31 (6.2%)
3+	6 (1.6%)	0 (0.0%)	29 (35.8%)	35 (7.0%)
Surgery				
None	368 (95.3%)	31 (100.0%)	35 (43.2%)	434 (87.2%)
BM	0 (0.0%)	0 (0.0%)	2 (2.5%)	2 (0.4%)
BO	18 (4.7%)	0 (0.0%)	23 (28.4%)	41 (8.2%)
Both	0 (0.0%)	0 (0.0%)	21 (25.9%)	21 (4.2%)

BO for Bilateral Oophorectomy, BM for Bilateral Mastectomy
SD for Standard Deviation.

Table 5.4: Descriptive statistics for the probands of *BRCA2* positive families.

	Breast Cancer	Ovarian Cancer	Unaffected	Total
N (%)	301 (79.6%)	10 (2.7%)	67 (17.7%)	378
Study entry age				
mean (SD)	47.6 (10.9)	53.2 (11.6)	46.6 (11.1)	47.5 (11.0)
min, max	22.0, 84.4	37.0, 76.3	22.6, 82.3	22.0, 84.4
Event age				
mean (SD)	43.5 (9.3)	51.4 (11.8)	54.2 (12.8)	45.6 (10.9)
min, max	21.8, 74.0	35.0, 75.0	26.6, 86.3	21.8, 86.3
BRCA Mutation				
Carrier	301 (100.0%)	10 (100.0%)	67 (100.0%)	378 (100.0%)
Noncarrier	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Untested	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
# of screening				
0	198 (65.8%)	8 (80.0%)	10 (14.9%)	216 (57.1%)
1	88 (29.2%)	2 (20.0%)	22 (32.8%)	112 (29.6%)
2	13 (4.3%)	0 (0.0%)	8 (11.9%)	21 (5.6%)
3+	2 (0.7%)	0 (0.0%)	27 (40.3%)	29 (7.7%)
Surgery				
None	289 (96.0%)	10 (100.0%)	32 (47.8%)	331 (87.6%)
BM	0 (0.0%)	0 (0.0%)	1 (1.5%)	1 (0.3%)
BO	12 (4.0%)	0 (0.0%)	22 (32.8%)	34 (9.0%)
Both	0 (0.0%)	0 (0.0%)	12 (17.9%)	12 (3.2%)

BO for Bilateral Oophorectomy, BM for Bilateral Mastectomy
SD for Standard Deviation.

Table 5.5: Descriptive statistics for the Mammographic Screening and Bilateral Oophorectomy ages in years

	BRCA1	BRCA2
	Mean (SD)	
First MS age	40.6 (12.4)	42.6 (13.2)
BO age	45.7 (10.4)	46.9 (10.3)
Gaptime		
MS1 - MS2	9.0 (7.6)	9.2 (8.2)
MS2 - MS3	6.0 (4.3)	6.2 (4.9)
BO - MS	12.1 (10.2)	16.0 (14.7)
MS - BO	8.1 (6.9)	9.3 (8.3)

MS1,MS2 and MS3 stand for the first, second and third Mammogrphic Screenings, respectively, BO for Bilateral Oophorectomy, SD for Standard Deviation.

5.2 Model specification

5.2.1 Effects of bilateral oophorectomy and mammographic screenings

We fit the shared frailty competing risks model with TVC discussed in Chapter 3 considering ovarian cancer and death as the competing risks for breast cancer. The conditional cause-specific hazard functions for breast cancer, ovarian cancer and death, denoted as h_1, h_2, h_3, respectively, are expressed as follow:

$$
\begin{aligned}
h_1(t|\boldsymbol{X}, z_1) &= h_{01}(t)\exp\{\beta_{g,1}G + g_o(X_o(t,t_o)) + g_{s1}(X_{s1}(t,t_{s1})) + g_{s2}(X_{s2}(t,t_{s2})) + g_{s3}(X_{s3}(t,t_{s3}))\}z_1 \\
h_2(t|\boldsymbol{X}, z_2) &= h_{02}(t)\exp\{\beta_{g,2}G\}z_2 \\
h_3(t|\boldsymbol{X}, z_3) &= h_{03}(t)\exp\{\beta_{g,3}G\}z_3, \qquad (5.1)
\end{aligned}
$$

where $h_{0j}(t), j = 1,2,3$, are the Weibull baseline hazard functions for event j, G is the time-invarying covariate for binary mutation status, $\beta_{g,j}$ are the coefficient of G associated with event j. z_j are the cause-specific shared frailty variables for event j. $X_o(t,t_o)$ and $X_s(t,t_s)$ are the time varying covariates for BO and MS, respectively, with surgery age t_o and screening age t_s; $X_o(t) = 1$ if $t \geq t_o$, 0 otherwise and $X_s(t) = 1$ if $t \geq t_s$. Note that BO and screening are associated with the risk of BC but not with the other competing risks.

The effect of BO, $g_o(\cdot)$, is defined in three ways: PE, ED, and CO, as follows:

$$
g_o(X_o(t,t_o)) = \begin{cases} 0 & \text{if } t < t_o \text{ (PE,ED,CO)} \\ \beta_o & \text{if } t \geq t_o \text{ (PE)} \\ \beta_o \cdot \exp\big\{-\eta_o(t-t_o)\big\} & \text{if } t \geq t_o \text{ (ED)} \\ \beta_o \cdot \exp\big\{-\eta_o(t-t_o)\big\} + \eta_{o,0} & \text{if } t \geq t_o \text{ (CO)} \end{cases}
$$

where β_o is the effect of the BO, η_o is the decay rate parameter (ED and CO models) and

$\eta_{o,0}$ is the decay convergence parameter (CO model).

For screening effects, we include up to three screenings by 3 TVCs:
$X_{s1}(t,t_{s1})$, $X_{s2}(t,t_{s2})$, $X_{s3}(t,t_{s3})$, where t_{s1}, t_{s2}, t_{s3} are ages in years for the first, second and third screening, respectively, and their effects are defined in three ways: PE, ED, and CO, as follows:

$$g_{s1}(X_{s1}(t,t_{s1})) = \begin{cases} 0 & \text{if } t < t_{s1} \text{ (PE,ED,CO)} \\ \beta_{s1} & \text{if } t_{s1} \le t < t_{s2} \text{ (PE)} \\ \beta_{s1} \cdot \exp\{-\eta_{s1}(t-t_{s1})\} & \text{if } t_{s1} \le t < t_{s2} \text{ (ED)} \\ \beta_{s1} \cdot \exp\{-\eta_{s1}(t-t_{s1})\} + \eta_{s1,0} & \text{if } t_{s1} \le t < t_{s2} \text{ (CO)} \\ 0 & \text{if } t \ge t_{s2} \text{ (PE,ED,CO)}, \end{cases}$$

where the effect of the first screening, β_{s1} is present only until the second screening occurs (PE model) and it decays with rate η_{s1} (ED and CO models) and converges to $\eta_{s1,0}$ (CO model). Decay rate and convergence parameters associated with the first screening, η_{s1} and $\eta_{s1,0}$, are only defined if we assume ED and CO TVC functions. Effects of the second and third screenings, $g_{s2}(\cdot)$ and $g_{s3}(\cdot)$ are defined similarly:

$$g_{s2}(X_{s2}(t,t_{s2})) = \begin{cases} 0 & \text{if } t < t_{s2} \text{ (PE,ED,CO)} \\ \beta_{s2} & \text{if } t_{s2} \le t < t_{s3} \text{ (PE)} \\ \beta_{s2} \cdot \exp\{-\eta_{s2}(t-t_{s2})\} & \text{if } t_{s2} \le t < t_{s3} \text{ (ED)} \\ \beta_{s2} \cdot \exp\{-\eta_{s2}(t-t_{s2})\} + \eta_{s2,0} & \text{if } t_{s2} \le t < t_{s3} \text{ (CO)} \\ 0 & \text{if } t \ge t_{s3} \text{ (PE,ED,CO)}, \end{cases}$$

$$g_{s3}(X_{s3}(t,t_{s3})) = \begin{cases} 0 & \text{if } t < t_{s3} \text{ (PE,ED,CO)} \\ \beta_{s3} & \text{if } t \geq t_{s3} \text{ (PE)} \\ \beta_{s3} \cdot \exp\{-\eta_{s3}(t-t_{s3})\} & \text{if } t \geq t_{s3} \text{ (ED)} \\ \beta_{s3} \cdot \exp\{-\eta_{s3}(t-t_{s3})\} + \eta_{s3,0} & \text{if } t \geq t_{s3} \text{ (CO) .} \end{cases}$$

5.2.2 Missing mutation status

Binary mutation status variable is included in the model (5.1) as time invarying covariate.As shown in Tables 5.1 and 5.2, mutation status is missing for about 60% of the women, while most of the known mutation status are carriers ($\sim$73% in *BRCA1* and $\sim$70% in *BRCA2*). We impute missing mutation status based on empirical carrier probabilities from the observed data and familial relationship and employ the robust variance estimators as described in Section 3.4 for our proposed model with the imputed data.

Let G_{f_i} be the mutation status variable for family member i of the family f. We empirically calculate the mutation carrier probability, $\omega_{g_{f_i}}$ for individual i in family f from observed mutation status, G^o for each subset of the data which they share same disease status D, other covariate values S and familial relationships to the proband:

$$\omega_{g_{f_i}} = P(G_{f_i} = g_{f_i} | D, S, G^o),$$

where g_{f_i} can take the value of 1 or 0 to represent a carrier or non-carrier of the mutated gene. Then, we sample the carrier status based on this conditional carrier probability. Imputation process was done single time due to computational burden for making inference from multiple imputations. We note that parameter estimates in our model from five multiple imputations did not vary substantially.

5.3 Analysis results

Table 5.6 presents the analysis results of the *BRCA1* and *BRCA2* mutation positive families in terms of the maximum likelihood estimates (MLEs), robust standard errors (SE) and p-values for the model parameters. We chose the PE model for BO and ED for MS as it provided the lowest Alkaike information criterion (AIC) value among possible combinations of TVC functions.

Figures 5.2 to 5.9 present cause-specific penetrance estimates. Figures 5.2 (*BRCA1*) and 5.6 (*BRCA2*) show breast and ovarian cancer penetrances for carriers and non-carriers who have neither screenings nor BO. For *BRCA1*, Figures 5.3 and 5.4 show breast cancer penetrance given the single and three screenings and Figure 5.5 shows the breast cancer penetrance given BO. For *BRCA2*, Figures 5.7 and 5.8 show breast cancer penetrance given the single and three screenings and Figure 5.9 shows the breast penetrance given BO.

For the breast cancer penetrance estimation, we use $t_{s1} = 40$, $t_{s2} = 45$, $t_{s3} = 50$, $t_o = 45$, which are close to the mean ages found in Table 5.5.

5.3.1 Relative risk

Log relative risks of the *BRCA* mutation, screenings and BO are presented in Table 5.6. Log relative risks of the mutation carrier towards breast cancer are $\hat{\beta}_{g1} = 2.25$ ($\widehat{\text{SE}} = 0.23$) in *BRCA1* and $\hat{\beta}_{g1} = 1.86$ ($\widehat{\text{SE}} = 0.20$) for *BRCA2*. The result indicates that being a mutation carrier increases the cause-specific hazard of developing breast cancer by approximately 9.49 times compared to non-carrier and 6.42 times in *BRCA2* adjusting for the screenings, BO and familial correlation.

Log relative risks of BO for breast cancer is $\hat{\beta}_o = -0.20$ ($\widehat{\text{SE}} = 0.20$) for *BRCA1* and $\hat{\beta}_o = -0.62$ ($\widehat{\text{SE}} = 0.22$) for *BRCA2*. BO decreases the cause-specific hazard of devel-

Table 5.6: Modelling results for *BRCA1/2* mutation positive families with a shared frailty competing risk model with Permanent Exposure (PE) TVC for Bilateral Oophorectomy (BO) and Exponential Decay (ED) for Mammographic Screenings (MS).

Shared frailty Competing Risk Model with PE BO and ED MS						
	BRCA1 families			BRCA2 families		
	MLE	SE	p-value	MLE	SE	p-value
λ_1	0.0093	0.0004	< 0.001	0.0098	0.0003	< 0.001
ρ_1	2.6041	0.0720	< 0.001	2.9918	0.0937	< 0.001
λ_2	0.0084	0.0005	< 0.001	0.0053	0.0010	< 0.001
ρ_2	3.1831	0.1435	< 0.001	2.3299	0.2902	< 0.001
λ_3	0.0153	0.0001	< 0.001	0.0154	0.0002	< 0.001
ρ_3	4.2225	0.1428	< 0.001	4.0877	0.1133	< 0.001
Breast Cancer						
β_{g1}	2.1786	0.1227	< 0.001	1.8606	0.2036	< 0.001
β_{s1}	3.3494	0.2907	< 0.001	3.8261	0.2556	< 0.001
β_{s2}	3.3548	0.4094	< 0.001	3.8273	0.4841	< 0.001
β_{s3}	2.8169	0.7050	< 0.001	1.1431	0.2661	< 0.001
β_o	-0.1971	0.1984	0.3204	-0.6191	0.2210	0.0051
η_{s1}	2.8946	0.8464	< 0.001	1.6931	0.1945	< 0.001
η_{s2}	4.1197	1.4184	0.0037	1.3482	0.5250	0.0102
η_{s3}	2.7508	1.7201	0.1098	0.0136	0.0084	0.0574
k_1	3.2566	0.2550	< 0.001	2.6625	0.5988	< 0.001
Ovarian Cancer						
β_{g2}	1.1757	0.2176	< 0.001	-0.9350	0.3432	0.0064
k_2	1.6882	0.2092	¡0.001	0.1993	0.0636	0.0017
Death						
β_{g3}	-0.4401	0.1515	0.0037	-0.6806	0.1158	< 0.001

MLE is Maximum Likelihood Estimates, SE is robust Standard Error.

oping breast cancer by approximately 0.76 times in *BRCA1* and 0.54 times in *BRCA2* adjusting for the screenings, mutation status and familial correlation.

Screenings drastically increase cause-specific hazard of developing breast cancer. Log relative risks of the first, second and third screenings are $\hat{\beta}_{s1} = 2.18$ ($\widehat{\text{SE}} = 0.12$), $\hat{\beta}_{s2} = 3.35$ ($\widehat{\text{SE}} = 0.29$), $\hat{\beta}_{s3} = 2.82$ ($\widehat{\text{SE}} = 0.71$) in *BRCA1* and $\hat{\beta}_{s1} = 3.83$ ($\widehat{\text{SE}} = 0.26$), $\hat{\beta}_{s2} = 3.83$ ($\widehat{\text{SE}} = 0.48$), $\hat{\beta}_{s3} = 1.14$ ($\widehat{\text{SE}} = 0.27$) in *BRCA2*. However, the effects of screenings decay with time at relatively high rates.

There is a positive association between *BRCA* mutation and the cause-specific hazard of developing ovarian cancer in *BRCA1*, $\hat{\beta}_{g2} = 1.18$ ($\hat{\text{SE}} = 0.22$). Being a mutation carrier increases the cause-specific hazard of developing ovarian cancer by approximately 3.25 times compared to non-carrier. In contrast, *BRCA2* mutation shows negative association with ovarian cancer $\hat{\beta}_{g2} = -0.94$ ($\hat{\text{SE}} = 0.34$).

The estimates of frailty parameters k_1 and k_2 are 3.26 (95% CI between 2.76 and 3.76) and 1.69 (95% CI between 1.28 and 2.10) in *BRCA1* and 2.66 (95% CI between 1.49 and 3.84) and 0.19 (95% CI between 0.075 and 0.32) in *BRCA2*. They correspond to Kendall's τ around 0.13 for breast cancer and 0.23 for ovarian cancer in *BRCA1* and 0.16 for breast cancer and 0.72 for ovarian cancer in *BRCA2*.

5.3.2 *BRCA1* penetrance estimations

Figure 5.2 shows the estimated cause-specific penetrance functions for breast cancer (left) and ovarian cancer (right) for those who have no screenings nor BO for *BRCA1* mutation positive families. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted lines are 95% confidence intervals. Breast cancer penetrance by age 70 is estimated around 60.3% (95% CI between 55.9% and 64.6%) for mutation carriers while for non-carriers, it is 11.6% (95% CI between 9.9% and 13.3%). Ovarian cancer penetrance at age 70 is much lower than breast cancer penetrance, 11.4% (95% CI between 9.0% and 13.9%) for carriers and 5.4% (95% CI between 4.2% and 6.6%) for non-carriers.

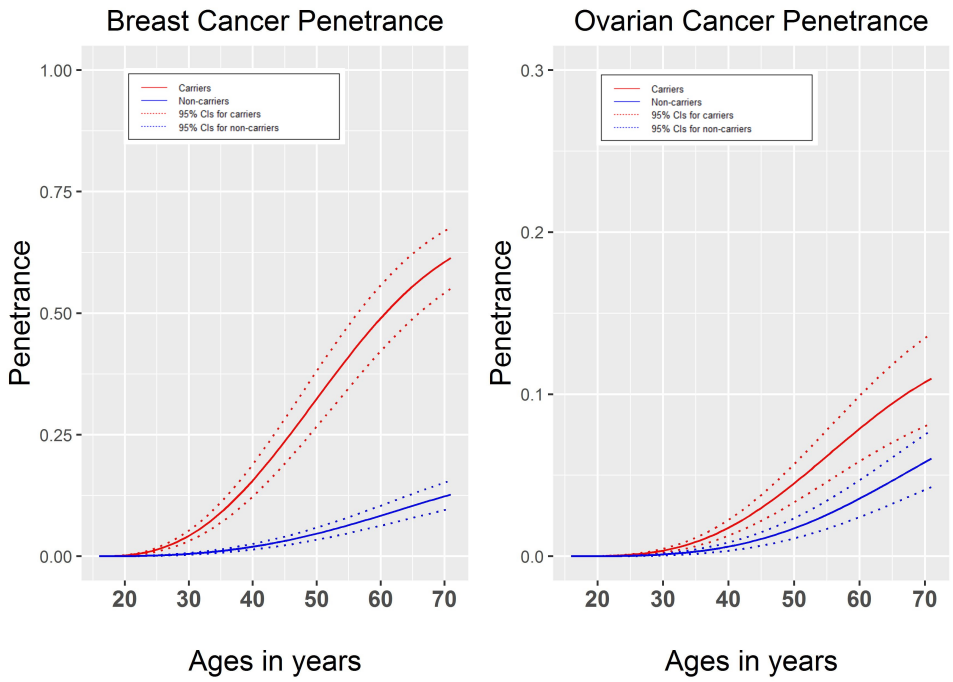

Figure 5.2: (*BRCA1*) Penetrance estimations for breast cancer (left panel) and ovarian cancer (right panel) conditioning no screening activities and surgeries. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals.

Figure 5.3 presents breast cancer penetrance estimates for those who have single screening and no BO from *BRCA1*. The left most plot is penetrance estimate for those without screening and is equivalent to the left plot of Figure 5.2. The remaining three plots to the right show penetrances given that the first screening occured at age 35, 40 and 45, respectively.

Given that the first screening occured at age 35, penetrance by age 70 is estimated around 67.4% (95% CI between 62.5% and 72.3%) for mutation carriers while for non-carriers, it is 12.3% (95% CI between 10.5% and 14.2%). With the first screening at age 45, penetrance by age 70 is estimated around 75.5% (95% CI between 69.5% and 81.5%) for mutation carriers and 12.8% (95% CI between 10.9% and 14.8%) for non-carriers. Penetrance of the non-carriers is not substantially increased by the age of screening, whereas penetrance of the carriers who have the first screening earlier in life have lower penetrance. This could be an indication that having a screening visit early is beneficial.

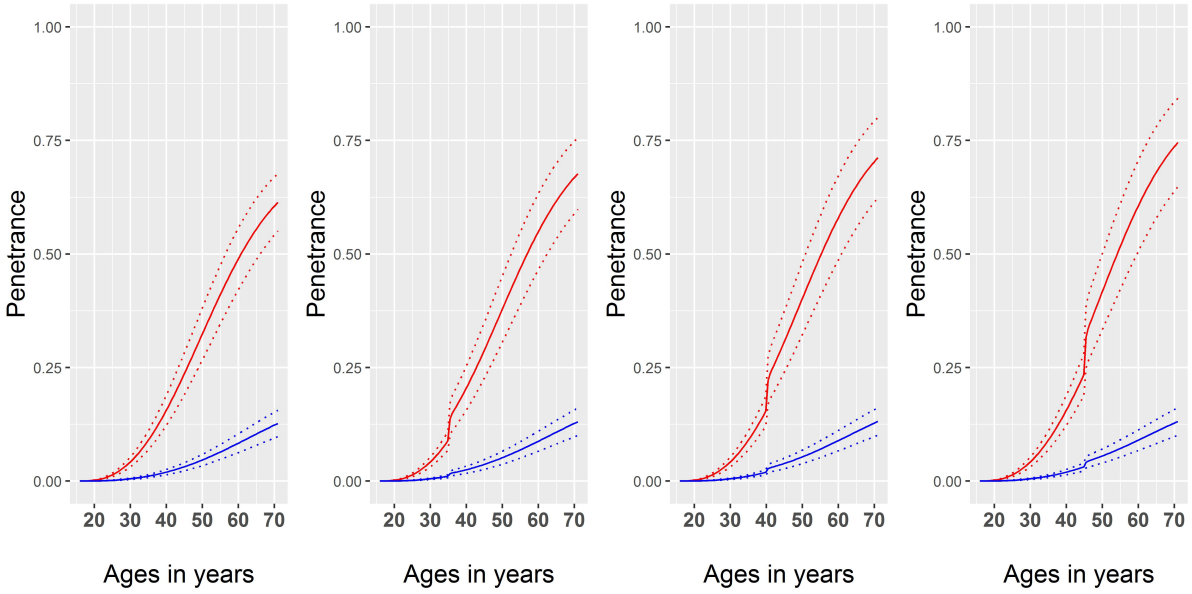

Figure 5.3: (*BRCA1*) Breast cancer (BC) penetrance estimations with the first mammographic screenings. The left most plot represents BC penetrance with no screening. To the right, they describe penetrance estimation with the first screening at age 35, 40 and 45, respectively. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals.

Figure 5.4 shows breast cancer penetrance for those who have three consecutive screenings for *BRCA1* positive families. The left most plot depicts the penetrance estimate without any screening. Plots in the middle and the right show penetrance estimates given that the first, second and third screening occured at ages 30, 35 and 40, and at ages 35, 40 and 45, respectively.

Similar to Figure 5.3, those who have the first screening early have lower penetrance compared to those lagged by 5 years. As well, penetrance is much higher if the subjects have multiple screenings (90.5% vs 67.4%) compared to those with single screening even though both groups have the first screen at age of 35.

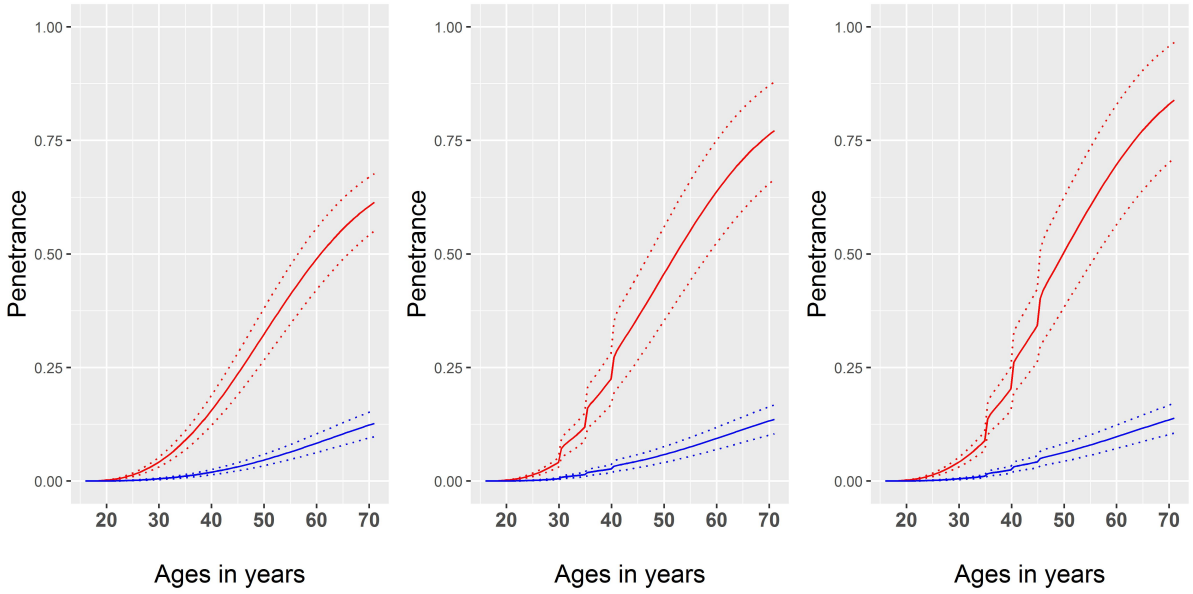

Figure 5.4: (*BRCA1*) Breast cancer (BC) penetrance estimations with three mammographic screenings. The left most plot represents BC penetrance with no screening. To the right, they describe penetrance estimations with the first screening at age 30 and 35 with the consecutive screening gap times of 5 years. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals.

Figure 5.5 presents breast cancer penetrance for subgroups who have BO for *BRCA1* positive families. The left most plot is the penetrance function estimated for those without BO, which is equivalent to the left plot in Figure 5.2. Remaining three plots to the right are penetrances given that BO occured at age 40, 45 and 50, respectively.

Given that BO occured at age 40, penetrance by age 70 is estimated around 64.9% (95% CI between 54.1% and 75.8%) for mutation carriers while for non-carriers, it is 9.5% (95% CI between 6.5% and 12.5%). Given that BO occured at age 45, penetrance by age 70 is estimated around 68.1% (95% CI between 57.5% and 78.7%) for the mutation carriers while for non-carriers, it is 9.6% (95% CI between 6.8% and 12.3%). Given that BO occured at age 50, penetrance by age 70 is estimated around 70.4% (95% CI between 60.3% and 80.5%) for the mutation carriers while for non-carriers, it is 9.7% (95% CI between 7.3% and 12.1%). Even though estimated BO effect, $\hat{\beta}_o = -0.20$, has negative but statistically insignificant association with cause-specific hazard of breast cancer, it turns out that the effect does not affect the penetrance in the same direction.

This is a well known property in cause-specific hazard approach. There is no one-to-one correspondence between hazard and penetrance functions.

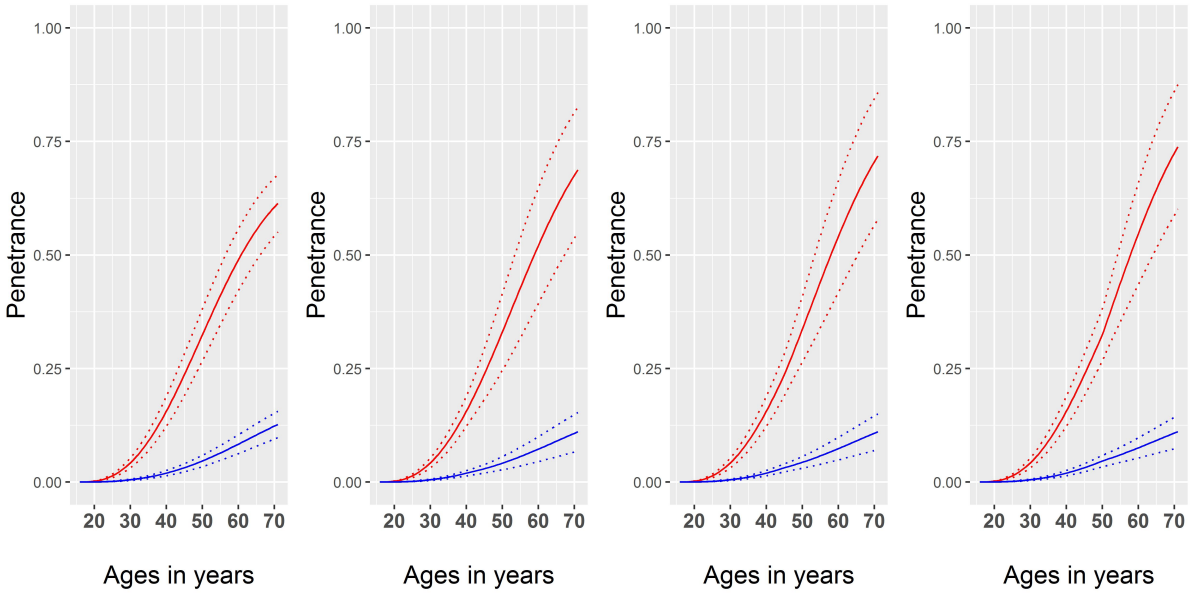

Figure 5.5: (*BRCA1*) Breast cancer (BC) penetrance estimations with bilateral oophorectomy. The left most plot represents BC penetrance with no BO. To the right, they describe penetrance estimation with the BO at age 40, 45 and 50. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals.

5.3.3 *BRCA2* penetrance estimations

Figure 5.6 shows the estimated cause-specific penetrance for breast cancer (left) and ovarian cancer (right) of the subgroups who do not have screenings and BO for *BRCA2* mutation positive families. Breast cancer penetrance by age 70 is estimated around 53.0% (95% CI between 48.0% and 57.9%) for mutation carriers while for non-carriers, it is 11.5% (95% CI between 9.6% and 13.4%). Ovarian cancer penetrance at age 70 is much lower than breast cancer penetrance, 1.4% (95% CI between 0.6% and 2.2%) for carriers and 4.0% (95% CI between 2.6% and 5.4%) for non-carriers. It is quite peculiar that the carrier exhibits lower penetrance than non-carriers and further investigation is required.

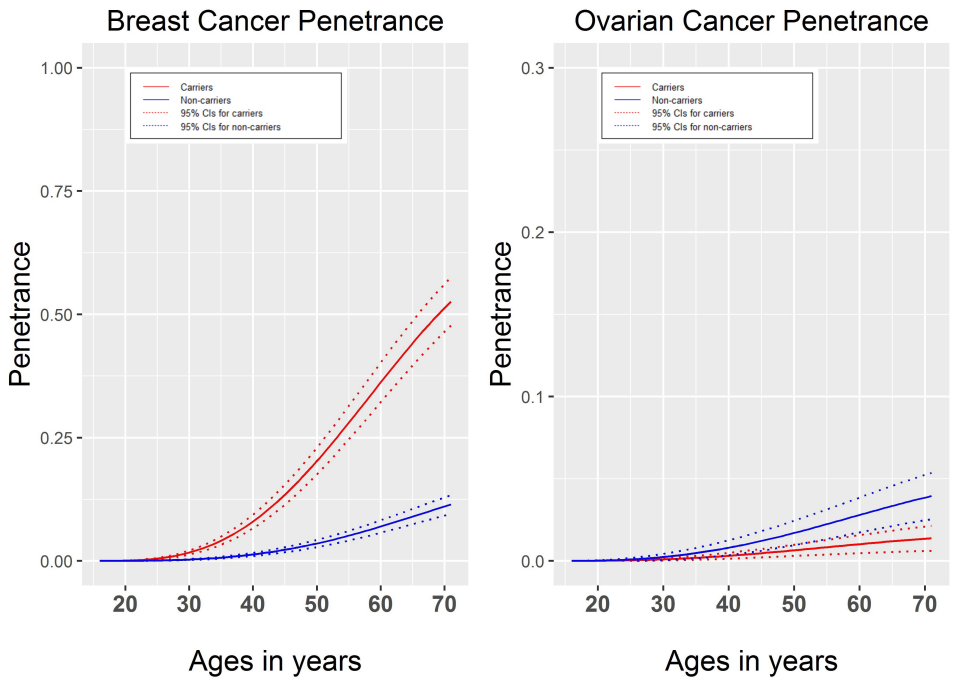

Figure 5.6: (*BRCA2*) Penetrance estimations for breast cancer (left panel) and ovarian cancer (right panel) conditioning no screening activities and surgeries. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals.

Figure 5.7 presents breast cancer penetrance estimates for subgroups who have single screening and no BO for *BRCA2*. The left most plot is the penetrance function estimated for those without screening, which is equivalent to the left plot in Figure 5.6. Remaining three plots to the right show penetrances given that the first screening occured at age 35, 40 and 45, respectively.

Given that the first screening occured at age 35, penetrance by age 70 is estimated around 56.8% (95% CI between 51.5% and 62.1%) for mutation carriers while for non-carriers, it is 12.2% (95% CI between 10.1% and 14.3%). Given that the first screening occured at age 45, penetrance by age 70 is estimated around 62.2% (95% CI between 56.0% and 68.4%) for mutation carriers while for non-carriers, it is 12.8% (95% CI between 10.6% and 15.0%).

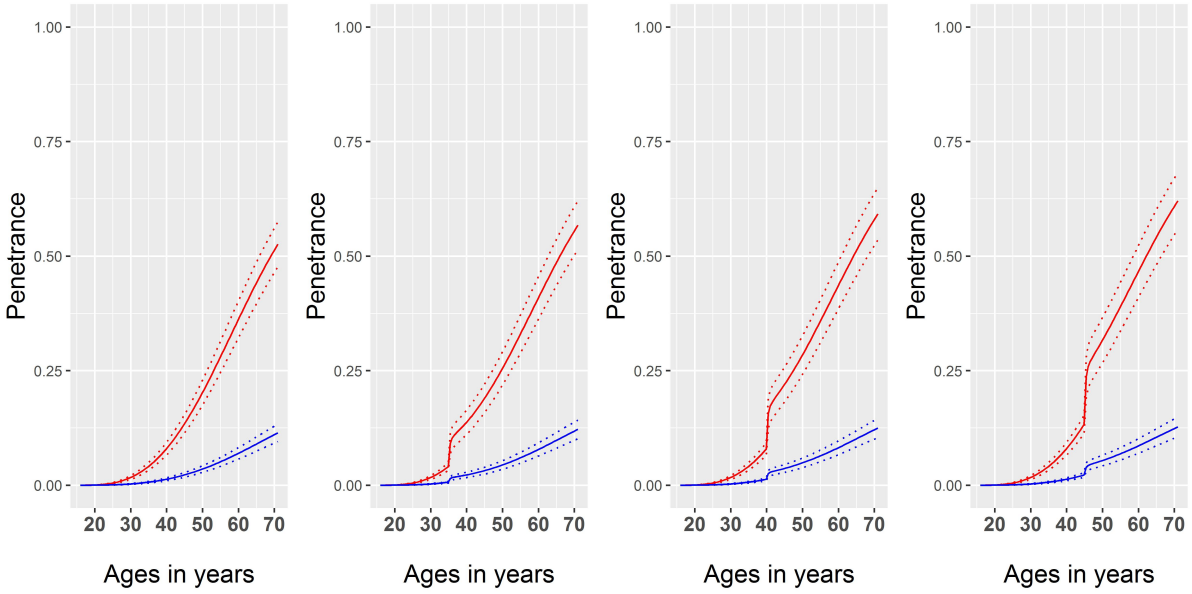

Figure 5.7: (*BRCA2*) Breast cancer (BC) penetrance estimations with the first mammographic screenings. The left most plot represents BC penetrance with no screening. To the right, they describe penetrance estimation with the first screening at age 35, 40 and 45, respectively. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals.

Figure 5.8 shows breast cancer penetrance for those who have three consecutive screenings for *BRCA2* positive families. The left most plot is the penetrance function estimated for those without any screening, which is equivalent to the left plot in Figure 5.6. Plots in the middle and the right show penetrance estimates given that the first, second and third screening occured at ages 30, 35 and 40, and at ages 35, 40 and 45, respectively.

Given that the first, second and third screenings occured at age 30,35 and 40, the penetrance by age 70 is estimated around 88.2% (95% CI between 74.0% and 88.6%) for mutation carriers while for non-carriers, it is 13.4% (95% CI between 11.3% and 15.6%). Given that the first, second and third screenings occured at age 35, 40 and 45, penetrance by age 70 is estimated around 90.5% (95% CI between 81.6% and 99.5%) for mutation carriers while for non-carriers, it is 14.1% (95% CI between 11.7% and 16.4%).

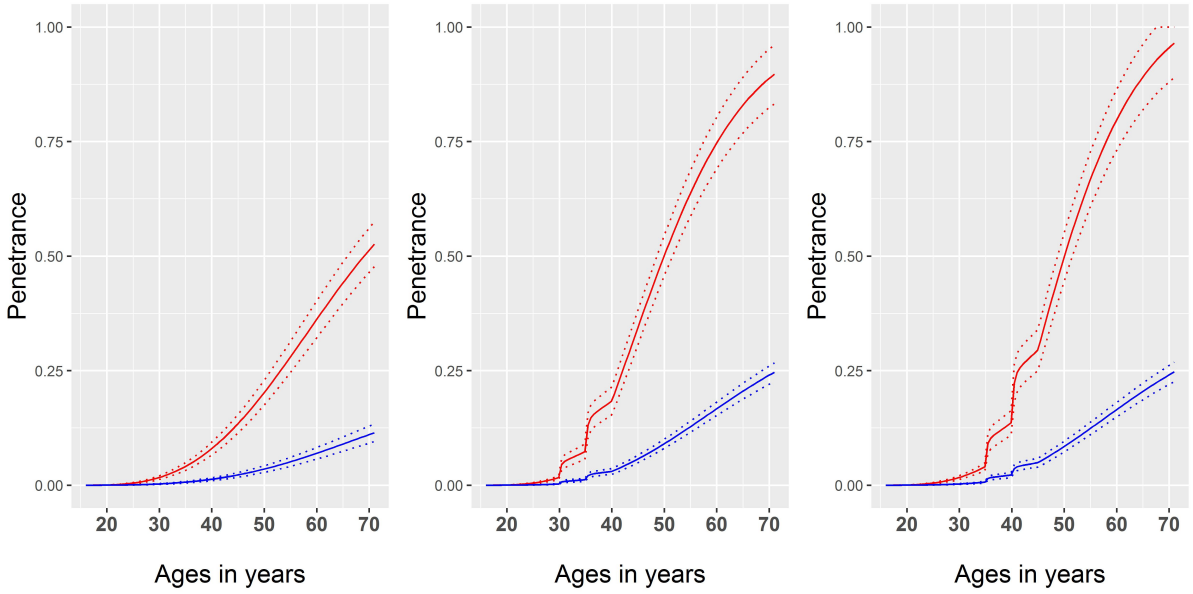

Figure 5.8: (*BRCA2*) Breast cancer (BC) penetrance estimations with three mammographic screenings. The left most plot represents BC penetrance with no screening. To the right, they describe penetrance estimations with the first screening at age 30 and 35 with the consecutive screening gap times of 5 years. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals.

Figure 5.9 presents breast cancer penetrance for subgroups who have BO for *BRCA2* positive families. The left most plot is estimate without BO and equivalent to the left plot in Figure 5.2. Remaining three plots to the right are penetrances given that BO occured at age 40, 45 and 50, respectively.

In *BRCA2*, BO decreases the cause-specific penetrance by age 70 contrary to *BRCA1*. Given that BO occured at age 40, penetrance by age 70 is estimated around 40.1% (95% CI between 29.5% and 50.8%) for mutation carriers while for non-carriers, it is 7.0% (95% CI between 4.4% and 9.7%). Given that BO occured at age 50, penetrance by age 70 is estimated around 45.2% (95% CI between 35.9% and 54.6%) for the mutation carriers while for non-carriers, it is 7.7% (95% CI between 5.5% and 9.9%). Similarly to the screening, earlier BO is beneficial since the life-time risk of breast cancer is reduced by about 5%.

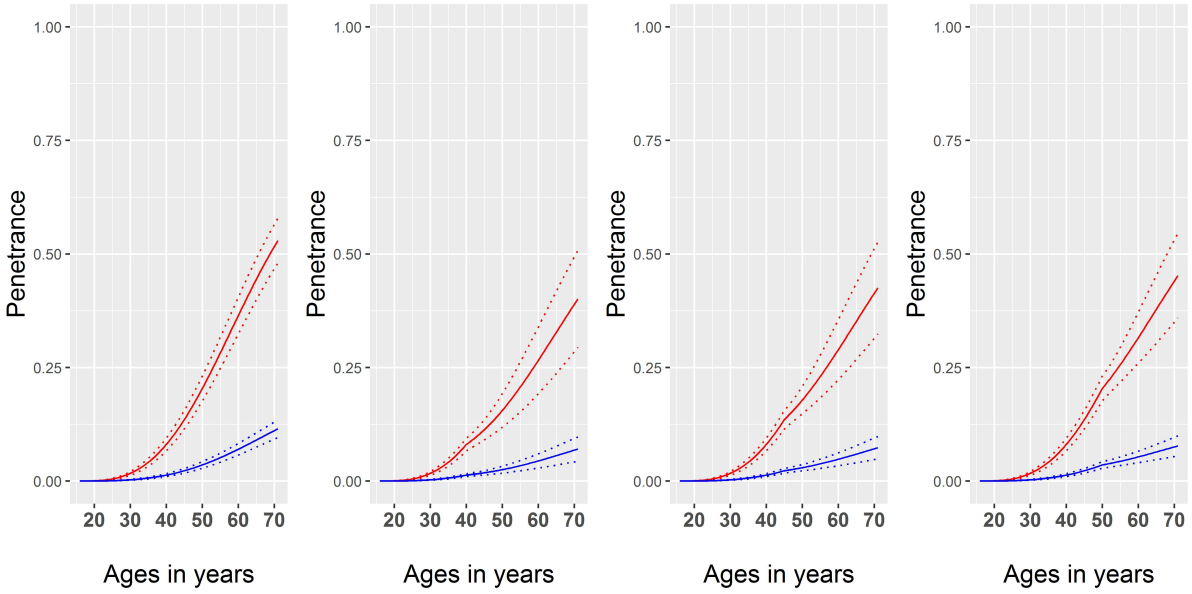

Figure 5.9: (*BRCA2*) Breast cancer (BC) penetrance estimations with bilateral oophorectomy. The left most plot represents BC penetrance with no BO. To the right, they describe penetrance estimation with the BO at age 40, 45 and 50. Solid red lines are penetrance estimates for mutation carriers and blue lines are for non-carriers. Dotted line represents 95% confidence intervals.

5.3.4 Summary

In both *BRCA1* and *BRCA2* families, there are statistically significant associations between mutation and cause-specific hazard of each cancer ($p < 0.001$). Being mutation carrier increases the cause-specific penetrance for breast cancer in *BRCA1/2*. Ovarian cancer penetrance is also affected by mutation status but in an opposite direction for *BRCA1/2*. BO has statistically significant negative association with cause-specific hazard of breast cancer in *BRCA2*. BO decreases the breast cancer penetrance by age at 70 (95% CI 0.48-0.58 for non-BO vs 0.29-0.51 for BO). However, it slightly increases penetrance in *BRCA1* (95% CI 0.56-0.64 for non-BO vs 0.54-0.76 for BO). MS is associated with the increase in both cause-specific hazard and penetrance of breast cancer possibly due to the reason that MS detects cancer. However, their effects decay to zero very fast. Failure times for ovarian cancer have stronger familial correlations than those of breast cancer.

Chapter 6

Discussion

6.1 Summary

This thesis proposes a shared frailty model extended to competing risks framework with time varying covariates for data arising from family studies. A shared frailty model is specified to account for familial residual dependence and an ascertainment-corrected likelihood is used to correct selection bias in family studies. We also provide the cause-specific cumulative incidence function that estimates age-specific risks of disease with time-varying covariates. Application of the model is not limited to the illustrative study on BRCA mutation families. Rather, it can be applied to any study when one is concerned with clustered failure times having multiple endpoints along with time varying binary covariates. However, further work is required to incorporate time dependent continuous or categorical variables.

We evaluated the performance of the estimators for the parameters in the model and the plug-in estimator of the penetrance function via simulations. The simulation results show that model parameters and penetrance estimators work well regardless of the number of families, degrees of familial dependence and the types of time varying covariate

functions. However, with $n = 250$ families, the model with the CO TVC exhibited convergence issues in about 5% of the total simulations. PE and ED TVC models did not have any convergence issues. This is possibly due to the increased number of parameters or the presence of multiple local maxima in partial likelihood function noted by Cox and Oakes (1984). Further investigation is required.

Analysis of the *BRCA* families shows that the mutation carrier is exposed to much higher risk of breast cancer and aligns with the previous studies. Risk of ovarian cancer is also elevated but in lesser degree than breast cancer. BO is associated with decrease in cause-specific hazard of breast cancer in *BRCA1/2* but there is no statistical significance in *BRCA1*. In *BRCA1*, this association does not correspond to the penetrance because penetrance by age 70 is slightly increased while both cause-specfic hazard and penetrance decrease with BO in *BRCA2*. We did not find any improvement in terms of AIC when we considered BO as ED and CO compared to PE. Their decay rate parameters converge to very small value indicating that the effect does not change much over time. When we handle the BO as TIC, it has much stronger association with breast cancer. On the other hand, the estimated effect of MS varies substantially with the choice of TVC function. Interestingly, when MS is handled as TIC, there is a statistically significant protective effect against breast cancer. MS as TVC unveils totally different association with breast cancer as it substantially increases the cause-specific hazard and penetrance but this effect mostly diminishes within about a year. Finally, failure times for ovarian cancer have stronger familial correlations than those of breast cancer.

6.2 Limitation and further work

Our analysis is limited to the first breast cancer in the presence of competing risks. However, successive events after the first event are also of interest in analysis of cancer cohorts. For instance, in evaluation of the screenings efficiency, one might be interested in the protective effect of the screenings leading to lower mortality rate after the first

breast cancer detection. Also, mortality rates could be different between cohorts who are affected by the breast and ovarian cancer. We have only considered the parametric approach in modelling the baseline hazard but in practice, use of a piece-wise constant baseline hazard function could be more desirable in some of the applications providing more flexible assumptions on baseline hazard.

The competing risk framework is not only limited to the cause-specific hazard approach. Subdistribution hazard and mixture model have their own advantages in terms of interpretability. For modeling the familial dependence, we only considered the independent frailty variables for the breast and ovarian cancer, but correlated frailties could better account for more complex dependence structure. In addtion, indepedent frailties assumption imposed in the analysis of Chapter 5 should be validated by assessing the correlated frailty model. Alternatively, a copula approach could also be considered beside frailty model.

In analysis of *BRCA* families, we treated the MS and BO as simple binary TVC assuming no measurement error and relaxed assumption on internal TVC. This is an important assumption when incorporating the TVC to the Cox model. Kalbfleisch and Prentice (2002) classified external and internal TVC. The former is often referred as exogeneous TVC since its path over time is not dependent on the subjects such as weather condition or predetermined drug dosage. More formally, the exogeneity condition states that the external TVC is a predictable process and the occurrence of the failure at time t does not affect the value of TVC for any time $s \geq t$ (Rizopoulos, 2012). In contrast, internal or endogeneous TVC is the measurement taken on the subjects. It complicates the statistical analysis since its path is affected by the occurence of the failure. For example, when the failure is defined as the death of the subject, then the existence of the TVC before infinitesimal time t ensures the survival probability of the subject at t to be 1 and failure time at t corresponds to the nonexistence of the TVC at $s \geq t$. Thus, the hazard function can be only defined up to time t, the existence time of the covariate process,

but not further, and the corresponding survival function is defined up to time t as well. We assume that the violation of the exogeneity condition is minimal for the MS and BO because their paths are not directly related to the event of interest.

In addition, we did not include other environmental risk factors potentially associated with breast cancer in women anticipating they have the minimal effects on the study results. Laden and Hunter (1998) investigated four potential risk factors in US women population: ionizing radiation, organochlorines, electromotive forces and smoking. They found that associations between the four pollutants and breast cancer are rather inconclusive and are substantially weaker compared to those of genetic factors.

Further work should involve more complex modelling when one is interested in incorporating such repeated measurements of longitudinal data. Potential candidates of those measurements could be the number of circulating tumor cells, immune response to a vaccine or a genetic biomarker. They are susceptible to measurement error and more difficult to relax the internal TVC assumption. Then, a linear mixed model can be considered for the longitudinal data. Joint modelling is a method to simultaneously model longitudinial data and the failure times not only accounting for measurement error but also to account for any association between longitudinal data and failure times (Ibrahim et al., 2010).

Appendix A

Additional simulation results

Table A.1: (1000 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Permanent Exposure TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations.

	Permanent Exposure Model														
	True	$k_1 = 7$, $\tau = 0.07$				True	$k_1 = 3.5$, $\tau = 0.13$				True	$k_1 = 1$, $\tau = 0.33$			
	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP
$log(\lambda_1)$	-4.83	-0.01	0.04	0.04	0.95	-4.83	0.00	0.04	0.04	0.95	-4.83	0.00	0.04	0.04	0.96
$log(\rho_1)$	0.88	0.00	0.02	0.02	0.96	0.88	0.00	0.02	0.02	0.95	0.88	0.00	0.02	0.02	0.97
$log(\lambda_2)$	-4.96	0.00	0.06	0.07	0.96	-4.96	-0.01	0.07	0.07	0.96	-4.96	0.00	0.07	0.07	0.96
$log(\rho_2)$	1.12	0.00	0.05	0.05	0.95	1.12	0.00	0.05	0.05	0.97	1.12	0.00	0.05	0.05	0.95
β_s	0.67	0.00	0.08	0.08	0.96	0.67	0.00	0.08	0.08	0.94	0.67	0.00	0.08	0.08	0.94
β_{g1}	1.95	0.01	0.08	0.08	0.94	1.95	0.00	0.08	0.08	0.94	1.95	0.00	0.08	0.08	0.94
β_{g2}	1.19	0.00	0.15	0.16	0.97	1.19	0.01	0.16	0.16	0.95	1.19	0.01	0.16	0.17	0.96
$log(k_1)$	1.95	0.12	0.62	0.58	0.95	1.25	0.05	0.36	0.32	0.96	0.00	0.01	0.17	0.18	0.96
$log(k_2)$	1.06	0.48	1.95	0.98	0.88	1.06	0.38	1.38	1.01	0.91	1.06	0.54	2.15	1.00	0.88
Cause-specific penetrance (%) given $t_{sc} = 35$															
Cause 1															
$F_1(40;S=0,G=0)$	2.06	-0.01	0.19	0.19	0.94	2.05	-0.01	0.20	0.19	0.94	2.04	0.00	0.19	0.20	0.95
$F_1(40;S=1,G=0)$	2.85	0.00	0.26	0.25	0.93	2.84	-0.01	0.26	0.25	0.94	2.81	0.00	0.27	0.27	0.93
$F_1(40;S=0,G=1)$	13.43	0.02	0.73	0.75	0.95	13.31	-0.02	0.81	0.79	0.94	12.72	0.00	0.89	0.92	0.95
$F_1(40;S=1,G=1)$	18.06	0.07	1.01	1.03	0.95	17.83	-0.02	1.09	1.08	0.95	16.80	0.01	1.21	1.23	0.96
$F_1(50;S=0,G=0)$	4.53	-0.02	0.39	0.37	0.93	4.52	-0.01	0.40	0.38	0.94	4.45	-0.01	0.39	0.40	0.96
$F_1(50;S=1,G=0)$	7.56	-0.01	0.68	0.65	0.95	7.52	-0.02	0.68	0.66	0.94	7.32	-0.01	0.69	0.70	0.94
$F_1(50;S=0,G=1)$	27.07	0.00	1.27	1.31	0.96	26.56	-0.01	1.42	1.39	0.95	24.38	0.01	1.59	1.63	0.95
$F_1(50;S=1,G=1)$	40.70	0.10	2.13	2.18	0.96	39.60	-0.03	2.26	2.27	0.94	35.11	0.02	2.40	2.44	0.95
$F_1(60;S=0,G=0)$	8.07	-0.05	0.66	0.64	0.93	8.02	-0.01	0.68	0.65	0.95	7.80	-0.02	0.66	0.68	0.96
$F_1(60;S=1,G=0)$	14.05	-0.04	1.20	1.17	0.94	13.91	-0.03	1.21	1.18	0.93	13.26	-0.03	1.20	1.21	0.94
$F_1(60;S=0,G=1)$	42.42	-0.04	1.81	1.87	0.96	41.23	0.01	2.01	1.99	0.95	36.41	0.03	2.23	2.30	0.95
$F_1(60;S=1,G=1)$	61.28	0.04	2.67	2.74	0.95	58.99	-0.04	2.85	2.88	0.94	50.29	0.03	3.06	3.14	0.95
$F_1(70;S=0,G=0)$	12.56	-0.09	0.99	0.97	0.94	12.45	-0.01	1.02	0.98	0.95	11.93	-0.02	0.99	1.02	0.95
$F_1(70;S=1,G=0)$	21.92	-0.08	1.78	1.74	0.95	21.58	-0.03	1.79	1.76	0.94	20.09	-0.04	1.73	1.76	0.95
$F_1(70;S=0,G=1)$	56.52	-0.09	2.19	2.27	0.96	54.51	0.03	2.41	2.42	0.96	46.80	0.03	2.67	2.77	0.96
$F_1(70;S=1,G=1)$	75.63	-0.04	2.57	2.66	0.95	72.59	-0.06	2.84	2.88	0.95	61.08	0.03	3.27	3.38	0.96
Cause 2															
$F_2(40;S=0,G=0)$	0.46	0.00	0.07	0.08	0.94	0.46	0.00	0.07	0.08	0.96	0.46	0.00	0.07	0.07	0.94
$F_2(40;S=1,G=0)$	0.46	0.00	0.07	0.07	0.94	0.46	0.00	0.07	0.07	0.96	0.46	0.00	0.07	0.07	0.94
$F_2(40;S=0,G=1)$	1.40	0.00	0.15	0.15	0.95	1.41	0.01	0.15	0.15	0.96	1.41	0.00	0.15	0.16	0.96
$F_2(40;S=1,G=1)$	1.38	0.00	0.15	0.15	0.95	1.38	0.01	0.15	0.15	0.96	1.39	0.00	0.15	0.15	0.96
$F_2(50;S=0,G=0)$	1.27	0.01	0.17	0.18	0.95	1.27	-0.01	0.17	0.18	0.95	1.27	0.00	0.17	0.18	0.94
$F_2(50;S=1,G=0)$	1.25	0.01	0.17	0.17	0.95	1.25	-0.01	0.17	0.17	0.95	1.25	0.00	0.17	0.17	0.95
$F_2(50;S=0,G=1)$	3.53	0.00	0.31	0.30	0.94	3.54	0.01	0.30	0.30	0.95	3.58	-0.01	0.31	0.32	0.96
$F_2(50;S=1,G=1)$	3.21	0.00	0.28	0.27	0.94	3.23	0.01	0.28	0.27	0.95	3.31	-0.01	0.28	0.29	0.96
$F_2(60;S=0,G=0)$	2.67	0.01	0.33	0.34	0.96	2.67	-0.01	0.34	0.34	0.95	2.67	-0.01	0.34	0.35	0.94
$F_2(60;S=1,G=0)$	2.57	0.01	0.32	0.33	0.96	2.57	-0.01	0.33	0.33	0.95	2.58	-0.01	0.33	0.34	0.94
$F_2(60;S=0,G=1)$	6.48	0.01	0.52	0.52	0.95	6.53	0.01	0.51	0.53	0.95	6.74	0.00	0.55	0.56	0.96
$F_2(60;S=1,G=1)$	5.32	0.00	0.43	0.43	0.95	5.42	0.01	0.43	0.44	0.95	5.83	-0.01	0.47	0.47	0.95
$F_2(70;S=0,G=0)$	4.73	0.03	0.58	0.61	0.96	4.73	-0.02	0.60	0.61	0.96	4.74	0.00	0.61	0.62	0.94
$F_2(70;S=1,G=0)$	4.45	0.02	0.55	0.57	0.95	4.45	-0.02	0.57	0.57	0.95	4.49	0.00	0.58	0.59	0.94
$F_2(70;S=0,G=1)$	9.68	0.02	0.80	0.82	0.95	9.85	0.02	0.77	0.83	0.97	10.52	0.01	0.88	0.91	0.96
$F_2(70;S=1,G=1)$	7.12	0.00	0.61	0.63	0.96	7.42	0.02	0.62	0.65	0.96	8.56	0.00	0.72	0.73	0.95

Table A.2: (1000 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Exponential Decay TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations.

Exponential Decay Model															
	True	$k_1 = 7$, $\tau = 0.07$				True	$k_1 = 3.5$, $\tau = 0.13$				True	$k_1 = 1$, $\tau = 0.33$			
	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP
$log(\lambda_1)$	-4.83	0.00	0.04	0.04	0.95	-4.83	0.00	0.04	0.04	0.95	-4.83	0.00	0.04	0.04	0.95
$log(\rho_1)$	0.83	0.00	0.02	0.02	0.95	0.83	0.00	0.02	0.02	0.95	0.83	0.00	0.02	0.02	0.95
$log(\lambda_2)$	-4.96	-0.01	0.06	0.07	0.96	-4.96	0.00	0.06	0.07	0.96	-4.96	0.00	0.06	0.07	0.98
$log(\rho_2)$	1.08	0.00	0.04	0.05	0.96	1.08	0.00	0.04	0.05	0.96	1.08	0.00	0.04	0.04	0.96
β_s	1.87	0.01	0.18	0.18	0.94	1.87	0.01	0.16	0.18	0.96	1.87	0.02	0.17	0.17	0.95
β_{g1}	1.86	0.00	0.08	0.08	0.96	1.86	0.00	0.08	0.08	0.96	1.86	0.01	0.08	0.08	0.96
β_{g2}	1.22	0.00	0.15	0.15	0.94	1.22	0.02	0.14	0.15	0.97	1.22	0.01	0.15	0.15	0.96
$log(k_1)$	1.95	0.16	0.91	0.59	0.94	1.25	0.04	0.32	0.32	0.95	0.00	0.02	0.16	0.17	0.97
$log(k_2)$	1.18	0.34	1.18	0.92	0.91	1.18	0.36	1.58	0.97	0.90	1.18	0.41	1.37	1.04	0.92
$log(\eta)$	-1.28	0.01	0.21	0.21	0.95	-1.28	0.00	0.20	0.22	0.96	-1.28	0.03	0.20	0.21	0.96
Cause-specific penetrance (%) given $t_{sc} = 35$															
Cause 1															
$F_1(40;S=0,G=0)$	2.43	0.00	0.20	0.21	0.96	2.42	0.00	0.21	0.21	0.95	2.40	0.02	0.22	0.22	0.95
$F_1(40;S=1,G=0)$	4.21	0.00	0.39	0.42	0.95	4.20	0.02	0.42	0.42	0.94	4.14	0.00	0.41	0.43	0.96
$F_1(40;S=0,G=1)$	14.36	-0.03	0.73	0.78	0.97	14.21	-0.01	0.79	0.82	0.96	13.55	0.14	0.93	0.94	0.95
$F_1(40;S=1,G=1)$	23.47	-0.05	1.53	1.61	0.96	23.09	0.02	1.59	1.63	0.96	21.41	0.10	1.63	1.68	0.95
$F_1(50;S=0,G=0)$	5.16	0.01	0.40	0.41	0.96	5.14	0.01	0.40	0.42	0.95	5.04	0.03	0.43	0.44	0.95
$F_1(50;S=1,G=0)$	7.30	0.02	0.61	0.66	0.95	7.26	0.05	0.66	0.67	0.95	7.08	0.01	0.65	0.68	0.96
$F_1(50;S=0,G=1)$	27.84	-0.06	1.21	1.31	0.97	27.30	-0.04	1.31	1.39	0.96	25.01	0.21	1.58	1.60	0.95
$F_1(50;S=1,G=1)$	36.87	-0.03	1.95	2.13	0.98	35.95	0.08	2.11	2.19	0.96	32.16	0.15	2.16	2.25	0.96
$F_1(60;S=0,G=0)$	8.91	0.01	0.66	0.68	0.96	8.85	0.01	0.65	0.69	0.95	8.59	0.04	0.70	0.72	0.95
$F_1(60;S=1,G=0)$	10.98	0.03	0.82	0.88	0.96	10.89	0.06	0.87	0.89	0.95	10.49	0.03	0.87	0.90	0.96
$F_1(60;S=0,G=1)$	42.49	-0.09	1.69	1.82	0.98	41.29	-0.06	1.82	1.93	0.96	36.44	0.26	2.14	2.19	0.95
$F_1(60;S=1,G=1)$	49.34	-0.03	2.05	2.26	0.98	47.76	0.06	2.28	2.36	0.95	41.55	0.22	2.45	2.55	0.95
$F_1(70;S=0,G=0)$	13.55	0.01	0.99	1.01	0.96	13.42	0.01	0.96	1.02	0.95	12.82	0.05	1.02	1.04	0.95
$F_1(70;S=1,G=0)$	15.49	0.03	1.09	1.16	0.96	15.32	0.06	1.13	1.17	0.95	14.54	0.04	1.15	1.18	0.95
$F_1(70;S=0,G=1)$	55.65	-0.10	2.03	2.19	0.98	53.68	-0.08	2.20	2.31	0.96	46.14	0.27	2.53	2.60	0.95
$F_1(70;S=1,G=1)$	60.49	-0.05	2.15	2.36	0.97	58.24	0.02	2.42	2.49	0.95	49.69	0.25	2.68	2.78	0.95
Cause 2															
$F_2(40;S=0,G=0)$	0.59	0.00	0.09	0.09	0.94	0.59	-0.01	0.08	0.09	0.96	0.59	0.00	0.09	0.09	0.94
$F_2(40;S=1,G=0)$	0.59	0.00	0.09	0.09	0.94	0.59	-0.01	0.08	0.09	0.96	0.59	0.00	0.09	0.09	0.94
$F_2(40;S=0,G=1)$	1.86	-0.01	0.18	0.18	0.95	1.86	0.00	0.18	0.18	0.94	1.86	0.01	0.18	0.19	0.96
$F_2(40;S=1,G=1)$	1.78	-0.01	0.17	0.17	0.95	1.78	0.00	0.17	0.17	0.94	1.80	0.01	0.17	0.18	0.96
$F_2(50;S=0,G=0)$	1.55	-0.01	0.19	0.20	0.94	1.55	-0.02	0.19	0.20	0.96	1.56	0.00	0.20	0.20	0.94
$F_2(50;S=1,G=0)$	1.52	-0.01	0.19	0.19	0.94	1.53	-0.02	0.18	0.19	0.96	1.53	0.00	0.19	0.20	0.94
$F_2(50;S=0,G=1)$	4.42	-0.02	0.34	0.34	0.95	4.44	0.01	0.34	0.34	0.95	4.49	0.01	0.36	0.36	0.96
$F_2(50;S=1,G=1)$	3.99	-0.02	0.31	0.30	0.95	4.02	0.01	0.31	0.31	0.95	4.12	0.01	0.33	0.32	0.95
$F_2(60;S=0,G=0)$	3.14	-0.02	0.36	0.38	0.95	3.14	-0.03	0.36	0.38	0.97	3.14	-0.01	0.37	0.39	0.95
$F_2(60;S=1,G=0)$	3.07	-0.02	0.35	0.37	0.95	3.07	-0.03	0.35	0.37	0.97	3.07	-0.01	0.36	0.38	0.95
$F_2(60;S=0,G=1)$	7.81	-0.04	0.56	0.57	0.95	7.87	0.03	0.57	0.58	0.97	8.13	0.02	0.63	0.62	0.94
$F_2(60;S=1,G=1)$	6.90	-0.03	0.50	0.50	0.96	6.99	0.01	0.51	0.51	0.96	7.38	0.02	0.56	0.55	0.95
$F_2(70;S=0,G=0)$	5.39	-0.03	0.60	0.64	0.95	5.39	-0.05	0.61	0.65	0.96	5.41	-0.01	0.62	0.66	0.96
$F_2(70;S=1,G=0)$	5.26	-0.03	0.58	0.63	0.95	5.26	-0.05	0.60	0.63	0.95	5.28	-0.01	0.60	0.65	0.96
$F_2(70;S=0,G=1)$	11.38	-0.04	0.82	0.86	0.96	11.57	0.04	0.86	0.88	0.96	12.34	0.03	0.98	0.96	0.94
$F_2(70;S=1,G=1)$	9.97	-0.04	0.74	0.76	0.96	10.22	0.02	0.77	0.79	0.95	11.20	0.03	0.88	0.87	0.94

Table A.3: (1000 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Cox and Oakes TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations.

	Cox and Oakes Model														
	True	$k_1 = 7$, $\tau = 0.07$				True	$k_1 = 3.5$, $\tau = 0.13$				True	$k_1 = 1$, $\tau = 0.33$			
	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP
$log(\lambda_1)$	-4.83	0.00	0.03	0.04	0.95	-4.83	0.00	0.03	0.04	0.98	-4.83	0.00	0.03	0.04	0.96
$log(\rho_1)$	0.83	0.00	0.02	0.02	0.95	0.83	0.00	0.02	0.02	0.98	0.83	0.00	0.02	0.02	0.96
$log(\lambda_2)$	-4.96	0.00	0.05	0.06	0.95	-4.96	-0.01	0.05	0.06	0.96	-4.96	0.00	0.05	0.06	0.97
$log(\rho_2)$	1.07	0.00	0.03	0.04	0.97	1.07	0.00	0.03	0.04	0.97	1.07	0.00	0.03	0.04	0.97
β_s	1.52	0.02	0.26	0.30	0.93	1.52	0.04	0.24	0.31	0.96	1.52	0.04	0.23	0.30	0.96
β_{g1}	2.08	0.01	0.07	0.08	0.97	2.08	0.01	0.07	0.08	0.95	2.08	0.00	0.07	0.08	0.95
β_{g2}	1.57	0.00	0.12	0.14	0.95	1.57	0.01	0.11	0.15	0.97	1.57	0.01	0.12	0.15	0.96
$log(k_1)$	1.95	0.03	0.43	0.51	0.94	1.25	0.04	0.25	0.30	0.97	0.00	0.00	0.13	0.16	0.97
$log(k_2)$	1.26	0.27	0.72	0.92	0.93	1.26	0.14	0.71	0.83	0.92	1.26	0.19	0.68	0.87	0.92
$log(\eta)$	-0.18	0.02	0.35	0.42	0.93	-0.18	0.03	0.33	0.43	0.94	-0.18	0.04	0.36	0.43	0.94
η_0	0.21	0.00	0.08	0.09	0.96	0.21	0.00	0.08	0.10	0.97	0.21	-0.01	0.08	0.10	0.96
Cause-specific penetrance (%) given $t_{sc} = 35$															
Cause 1															
$F_1(40;S=0,G=0)$	2.43	-0.02	0.20	0.21	0.94	2.42	-0.01	0.20	0.21	0.96	2.40	0.00	0.21	0.22	0.95
$F_1(40;S=1,G=0)$	3.25	-0.02	0.30	0.31	0.96	3.24	0.00	0.29	0.32	0.97	3.21	0.00	0.31	0.32	0.97
$F_1(40;S=0,G=1)$	17.45	-0.02	0.82	0.85	0.96	17.23	0.05	0.87	0.89	0.95	16.27	0.02	1.01	1.01	0.95
$F_1(40;S=1,G=1)$	22.47	0.02	1.40	1.42	0.95	22.12	0.12	1.35	1.45	0.96	20.56	0.02	1.41	1.48	0.96
$F_1(50;S=0,G=0)$	5.15	-0.04	0.40	0.40	0.95	5.13	-0.02	0.39	0.41	0.96	5.04	0.00	0.42	0.42	0.95
$F_1(50;S=1,G=0)$	6.57	-0.03	0.57	0.58	0.95	6.54	-0.01	0.54	0.58	0.96	6.39	0.01	0.57	0.59	0.96
$F_1(50;S=0,G=1)$	32.90	-0.03	1.30	1.38	0.96	32.16	0.10	1.43	1.45	0.95	29.06	0.04	1.63	1.65	0.96
$F_1(50;S=1,G=1)$	39.48	0.04	2.04	2.05	0.96	38.43	0.16	2.01	2.09	0.95	34.14	0.05	2.03	2.15	0.96
$F_1(60;S=0,G=0)$	8.91	-0.07	0.67	0.67	0.94	8.85	-0.03	0.63	0.68	0.96	8.58	0.01	0.68	0.69	0.95
$F_1(60;S=1,G=0)$	11.09	-0.06	0.94	0.96	0.95	11.00	-0.03	0.89	0.97	0.96	10.59	-0.01	0.92	0.97	0.96
$F_1(60;S=0,G=1)$	48.48	-0.06	1.71	1.84	0.96	46.94	0.15	1.88	1.93	0.95	40.89	0.05	2.12	2.16	0.95
$F_1(60;S=1,G=1)$	55.53	-0.04	2.51	2.57	0.95	53.57	0.14	2.50	2.61	0.95	46.03	-0.01	2.50	2.67	0.96
$F_1(70;S=0,G=0)$	13.54	-0.10	0.99	1.00	0.95	13.41	-0.04	0.93	1.01	0.97	12.81	0.01	0.98	1.01	0.96
$F_1(70;S=1,G=0)$	16.60	-0.11	1.40	1.43	0.94	16.41	-0.06	1.31	1.43	0.95	15.52	-0.02	1.33	1.40	0.96
$F_1(70;S=0,G=1)$	61.12	-0.10	1.93	2.09	0.96	58.82	0.18	2.11	2.19	0.95	50.11	0.05	2.38	2.46	0.95
$F_1(70;S=1,G=1)$	67.55	-0.13	2.58	2.65	0.95	64.90	0.11	2.60	2.73	0.95	54.88	-0.04	2.69	2.88	0.96
Cause 2															
$F_2(40;S=0,G=0)$	0.62	0.00	0.09	0.09	0.96	0.62	0.00	0.08	0.09	0.96	0.62	0.00	0.09	0.09	0.96
$F_2(40;S=1,G=0)$	0.62	0.00	0.08	0.09	0.96	0.62	0.00	0.08	0.09	0.96	0.62	0.00	0.09	0.09	0.96
$F_2(40;S=0,G=1)$	2.66	-0.01	0.22	0.23	0.96	2.67	0.00	0.22	0.23	0.95	2.68	0.01	0.23	0.24	0.96
$F_2(40;S=1,G=1)$	2.58	-0.01	0.22	0.22	0.96	2.59	0.00	0.21	0.22	0.95	2.61	0.01	0.22	0.23	0.96
$F_2(50;S=0,G=0)$	1.61	0.00	0.20	0.20	0.95	1.61	-0.01	0.19	0.20	0.96	1.61	0.00	0.20	0.21	0.96
$F_2(50;S=1,G=0)$	1.59	0.00	0.19	0.20	0.95	1.59	-0.01	0.18	0.20	0.96	1.59	0.00	0.20	0.20	0.95
$F_2(50;S=0,G=1)$	6.11	-0.02	0.42	0.42	0.95	6.13	-0.02	0.40	0.42	0.96	6.24	0.02	0.42	0.44	0.96
$F_2(50;S=1,G=1)$	5.66	-0.02	0.38	0.39	0.96	5.70	-0.03	0.37	0.39	0.95	5.86	0.01	0.39	0.41	0.96
$F_2(60;S=0,G=0)$	3.23	0.01	0.37	0.38	0.95	3.23	-0.03	0.35	0.38	0.96	3.24	0.00	0.37	0.39	0.96
$F_2(60;S=1,G=0)$	3.18	0.01	0.37	0.38	0.95	3.18	-0.03	0.34	0.37	0.96	3.18	0.00	0.37	0.38	0.96
$F_2(60;S=0,G=1)$	10.30	-0.01	0.68	0.67	0.95	10.42	-0.06	0.64	0.68	0.96	10.90	0.02	0.68	0.73	0.96
$F_2(60;S=1,G=1)$	9.22	-0.02	0.62	0.62	0.94	9.37	-0.07	0.60	0.62	0.95	10.00	0.02	0.62	0.67	0.97
$F_2(70;S=0,G=0)$	5.53	0.02	0.64	0.65	0.95	5.53	-0.06	0.59	0.65	0.96	5.55	0.00	0.63	0.67	0.97
$F_2(70;S=1,G=0)$	5.39	0.02	0.62	0.64	0.95	5.39	-0.06	0.58	0.63	0.96	5.42	0.00	0.61	0.65	0.97
$F_2(70;S=0,G=1)$	14.27	0.01	0.97	0.97	0.96	14.61	-0.12	0.93	0.99	0.97	15.91	0.02	1.00	1.10	0.97
$F_2(70;S=1,G=1)$	12.35	0.01	0.91	0.91	0.94	12.77	-0.11	0.90	0.93	0.94	14.36	0.02	0.94	1.02	0.97

Table A.4: (250 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Cox and Oakes TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations.

	Permanent Exposure Model														
	True	$k_1 = 7$, $\tau = 0.07$				True	$k_1 = 3.5$, $\tau = 0.13$				True	$k_1 = 1$, $\tau = 0.33$			
	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP
$log(\lambda_1)$	-4.83	0.00	0.08	0.08	0.95	-4.83	0.00	0.08	0.08	0.94	-4.83	0.00	0.08	0.08	0.94
$log(\rho_1)$	0.88	0.00	0.05	0.05	0.95	0.88	0.00	0.05	0.05	0.93	0.88	0.01	0.05	0.04	0.93
$log(\lambda_2)$	-4.96	-0.01	0.13	0.14	0.96	-4.96	0.00	0.14	0.14	0.94	-4.96	-0.01	0.13	0.14	0.95
$log(\rho_2)$	1.12	0.01	0.10	0.10	0.95	1.12	0.01	0.10	0.10	0.94	1.12	0.01	0.09	0.10	0.93
β_s	0.67	0.00	0.16	0.15	0.94	0.67	-0.01	0.15	0.16	0.96	0.67	-0.01	0.16	0.16	0.95
β_{g1}	1.95	-0.01	0.18	0.17	0.93	1.95	0.00	0.16	0.17	0.96	1.95	0.01	0.16	0.16	0.95
β_{g2}	1.19	0.03	0.31	0.33	0.96	1.19	0.02	0.33	0.33	0.96	1.19	0.04	0.33	0.34	0.96
$log(k_1)$	1.95	0.64	2.00	1.30	0.86	1.25	0.29	1.19	0.77	0.94	0.00	0.04	0.38	0.36	0.95
$log(k_2)$	1.06	0.85	2.77	1.63	0.79	1.06	0.84	2.18	1.61	0.78	1.06	0.90	2.73	1.68	0.77
Cause-specific penetrance (%) given $t_{sc} = 35$															
Cause 1															
$F_1(40;S=0,G=0)$	2.06	0.02	0.39	0.37	0.94	2.05	0.01	0.37	0.38	0.95	2.04	0.00	0.40	0.40	0.93
$F_1(40;S=1,G=0)$	2.85	0.04	0.52	0.50	0.93	2.84	0.01	0.49	0.51	0.94	2.81	-0.01	0.54	0.54	0.95
$F_1(40;S=0,G=1)$	13.43	-0.09	1.50	1.49	0.94	13.31	0.02	1.63	1.58	0.95	12.72	-0.01	1.86	1.84	0.96
$F_1(40;S=1,G=1)$	18.06	-0.03	2.08	2.05	0.94	17.83	0.00	2.20	2.16	0.94	16.80	-0.03	2.45	2.46	0.97
$F_1(50;S=0,G=0)$	4.53	0.04	0.78	0.74	0.94	4.52	0.03	0.75	0.76	0.95	4.45	0.02	0.81	0.81	0.94
$F_1(50;S=1,G=0)$	7.56	0.12	1.37	1.31	0.92	7.52	0.01	1.25	1.32	0.95	7.32	0.00	1.38	1.38	0.95
$F_1(50;S=0,G=1)$	27.07	-0.18	2.59	2.58	0.95	26.56	0.06	2.84	2.77	0.94	24.38	0.10	3.31	3.26	0.96
$F_1(50;S=1,G=1)$	40.70	-0.10	4.37	4.33	0.93	39.60	-0.06	4.51	4.51	0.94	35.11	-0.01	4.85	4.87	0.95
$F_1(60;S=0,G=0)$	8.07	0.07	1.31	1.27	0.94	8.02	0.06	1.28	1.30	0.95	7.80	0.06	1.37	1.38	0.95
$F_1(60;S=1,G=0)$	14.05	0.20	2.43	2.35	0.93	13.91	0.01	2.23	2.36	0.95	13.26	0.02	2.41	2.43	0.95
$F_1(60;S=0,G=1)$	42.42	-0.26	3.66	3.69	0.96	41.23	0.12	4.03	3.95	0.93	36.41	0.22	4.70	4.60	0.95
$F_1(60;S=1,G=1)$	61.28	-0.29	5.37	5.43	0.94	58.99	-0.14	5.68	5.71	0.94	50.29	0.00	6.24	6.25	0.95
$F_1(70;S=0,G=0)$	12.56	0.11	1.98	1.94	0.94	12.45	0.10	1.94	1.98	0.95	11.93	0.12	2.05	2.06	0.95
$F_1(70;S=1,G=0)$	21.92	0.27	3.60	3.50	0.94	21.58	0.01	3.31	3.50	0.95	20.09	0.07	3.50	3.53	0.95
$F_1(70;S=0,G=1)$	56.52	-0.34	4.40	4.47	0.95	54.51	0.14	4.87	4.79	0.92	46.80	0.29	5.70	5.55	0.94
$F_1(70;S=1,G=1)$	75.63	-0.45	5.11	5.26	0.94	72.59	-0.24	5.71	5.71	0.93	61.08	-0.03	6.74	6.72	0.94
Cause 2															
$F_2(40;S=0,G=0)$	0.46	-0.01	0.15	0.15	0.91	0.46	0.00	0.14	0.15	0.94	0.46	-0.01	0.15	0.15	0.93
$F_2(40;S=1,G=0)$	0.46	-0.01	0.14	0.15	0.91	0.46	0.00	0.14	0.15	0.94	0.46	-0.01	0.15	0.15	0.93
$F_2(40;S=0,G=1)$	1.40	-0.02	0.29	0.30	0.94	1.41	-0.01	0.31	0.30	0.94	1.41	-0.01	0.31	0.31	0.93
$F_2(40;S=1,G=1)$	1.38	-0.02	0.29	0.29	0.94	1.38	-0.01	0.30	0.29	0.94	1.39	-0.01	0.30	0.30	0.93
$F_2(50;S=0,G=0)$	1.27	-0.02	0.33	0.34	0.92	1.27	0.00	0.34	0.35	0.93	1.27	-0.02	0.35	0.35	0.94
$F_2(50;S=1,G=0)$	1.25	-0.02	0.33	0.34	0.92	1.25	0.00	0.33	0.34	0.93	1.25	-0.02	0.35	0.34	0.94
$F_2(50;S=0,G=1)$	3.53	-0.04	0.56	0.59	0.95	3.54	-0.02	0.60	0.60	0.94	3.58	-0.01	0.61	0.63	0.94
$F_2(50;S=1,G=1)$	3.21	-0.04	0.51	0.53	0.95	3.23	-0.02	0.55	0.54	0.94	3.31	-0.01	0.55	0.57	0.94
$F_2(60;S=0,G=0)$	2.67	-0.03	0.65	0.68	0.93	2.67	0.01	0.67	0.68	0.93	2.67	-0.04	0.69	0.69	0.94
$F_2(60;S=1,G=0)$	2.57	-0.03	0.62	0.65	0.94	2.57	0.01	0.65	0.66	0.94	2.58	-0.04	0.66	0.67	0.94
$F_2(60;S=0,G=1)$	6.48	-0.04	0.96	1.03	0.95	6.53	-0.01	1.02	1.05	0.95	6.74	0.01	1.08	1.12	0.96
$F_2(60;S=1,G=1)$	5.32	-0.04	0.81	0.85	0.95	5.42	-0.01	0.87	0.87	0.95	5.83	0.00	0.91	0.94	0.94
$F_2(70;S=0,G=0)$	4.73	-0.03	1.14	1.20	0.95	4.73	0.05	1.20	1.21	0.94	4.74	-0.04	1.20	1.23	0.93
$F_2(70;S=1,G=0)$	4.45	-0.04	1.07	1.13	0.94	4.45	0.05	1.13	1.14	0.93	4.49	-0.04	1.13	1.17	0.93
$F_2(70;S=0,G=1)$	9.68	-0.01	1.49	1.62	0.97	9.85	0.02	1.60	1.66	0.95	10.52	0.04	1.76	1.81	0.94
$F_2(70;S=1,G=1)$	7.12	-0.02	1.18	1.25	0.96	7.42	0.03	1.28	1.30	0.94	8.56	0.03	1.44	1.47	0.95

Table A.5: (250 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Cox and Oakes TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations.

Exponential Decay Model															
	True	$k_1 = 7$, $\tau = 0.07$				True	$k_1 = 3.5$, $\tau = 0.13$				True	$k_1 = 1$, $\tau = 0.33$			
	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP
$log(\lambda_1)$	-4.83	0.00	0.08	0.08	0.96	-4.83	-0.01	0.07	0.08	0.98	-4.83	-0.01	0.08	0.08	0.97
$log(\rho_1)$	0.83	0.00	0.04	0.05	0.97	0.83	0.00	0.04	0.05	0.96	0.83	0.00	0.04	0.04	0.95
$log(\lambda_2)$	-4.96	0.00	0.12	0.13	0.96	-4.96	-0.02	0.13	0.13	0.95	-4.96	-0.01	0.13	0.13	0.94
$log(\rho_2)$	1.08	0.01	0.09	0.09	0.95	1.08	0.00	0.09	0.09	0.95	1.08	0.00	0.09	0.09	0.94
β_s	1.87	0.02	0.37	0.36	0.93	1.87	0.02	0.34	0.36	0.95	1.87	0.03	0.33	0.34	0.95
β_{g1}	1.86	0.01	0.17	0.17	0.96	1.86	0.02	0.16	0.17	0.97	1.86	0.01	0.16	0.16	0.95
β_{g2}	1.22	0.02	0.30	0.30	0.95	1.22	0.03	0.30	0.30	0.95	1.22	0.04	0.31	0.31	0.95
$log(k_1)$	1.95	0.53	1.86	1.37	0.87	1.25	0.25	0.93	0.78	0.95	0.00	0.04	0.34	0.35	0.96
$log(k_2)$	1.18	0.64	2.16	1.45	0.81	1.18	0.88	2.56	1.62	0.79	1.18	0.78	2.19	1.94	0.80
$log(\eta)$	-1.28	0.00	0.47	0.45	0.92	-1.28	0.01	0.47	0.45	0.92	-1.28	-0.03	0.44	0.42	0.94
Cause-specific penetrance (%) given $t_{sc} = 35$															
Cause 1															
$F_1(40;S=0,G=0)$	2.43	-0.01	0.41	0.42	0.94	2.42	-0.02	0.41	0.43	0.94	2.40	0.01	0.43	0.45	0.95
$F_1(40;S=1,G=0)$	4.21	-0.02	0.82	0.83	0.93	4.20	-0.04	0.84	0.84	0.95	4.14	0.08	0.84	0.86	0.96
$F_1(40;S=0,G=1)$	14.36	-0.06	1.45	1.56	0.97	14.21	0.07	1.62	1.64	0.96	13.55	0.07	1.91	1.87	0.94
$F_1(40;S=1,G=1)$	23.47	-0.13	3.27	3.23	0.94	23.09	0.08	3.26	3.28	0.95	21.41	0.33	3.45	3.39	0.94
$F_1(50;S=0,G=0)$	5.16	-0.02	0.80	0.82	0.95	5.14	-0.05	0.79	0.83	0.95	5.04	0.02	0.84	0.87	0.96
$F_1(50;S=1,G=0)$	7.30	0.07	1.35	1.35	0.93	7.26	0.04	1.35	1.35	0.95	7.08	0.21	1.33	1.39	0.96
$F_1(50;S=0,G=1)$	27.84	-0.10	2.42	2.62	0.96	27.30	0.17	2.70	2.77	0.96	25.01	0.10	3.20	3.17	0.95
$F_1(50;S=1,G=1)$	36.87	0.22	4.30	4.38	0.96	35.95	0.54	4.38	4.40	0.95	32.16	0.70	4.56	4.58	0.95
$F_1(60;S=0,G=0)$	8.91	-0.04	1.33	1.35	0.95	8.85	-0.07	1.29	1.37	0.96	8.59	0.02	1.36	1.43	0.96
$F_1(60;S=1,G=0)$	10.98	0.12	1.82	1.82	0.94	10.89	0.08	1.78	1.81	0.96	10.49	0.26	1.76	1.86	0.97
$F_1(60;S=0,G=1)$	42.49	-0.15	3.36	3.64	0.97	41.29	0.28	3.71	3.86	0.96	36.44	0.13	4.32	4.36	0.95
$F_1(60;S=1,G=1)$	49.34	0.28	4.54	4.71	0.97	47.76	0.74	4.72	4.80	0.95	41.55	0.72	5.12	5.19	0.95
$F_1(70;S=0,G=0)$	13.55	-0.05	1.98	2.01	0.96	13.42	-0.08	1.89	2.03	0.96	12.82	0.02	1.97	2.08	0.97
$F_1(70;S=1,G=0)$	15.49	0.13	2.37	2.36	0.94	15.32	0.09	2.28	2.36	0.95	14.54	0.27	2.26	2.41	0.98
$F_1(70;S=0,G=1)$	55.65	-0.21	4.03	4.36	0.96	53.68	0.36	4.41	4.62	0.95	46.14	0.12	5.08	5.19	0.94
$F_1(70;S=1,G=1)$	60.49	0.17	4.61	4.83	0.95	58.24	0.73	4.89	5.03	0.94	49.69	0.61	5.51	5.62	0.94
Cause 2															
$F_2(40;S=0,G=0)$	0.59	0.00	0.18	0.17	0.92	0.59	0.00	0.17	0.17	0.94	0.59	-0.01	0.18	0.18	0.92
$F_2(40;S=1,G=0)$	0.59	0.00	0.18	0.17	0.92	0.59	0.00	0.17	0.17	0.94	0.59	-0.01	0.18	0.17	0.92
$F_2(40;S=0,G=1)$	1.86	-0.03	0.35	0.36	0.94	1.86	0.01	0.38	0.36	0.93	1.86	0.00	0.38	0.38	0.93
$F_2(40;S=1,G=1)$	1.78	-0.03	0.34	0.34	0.93	1.78	0.00	0.37	0.35	0.93	1.80	0.00	0.36	0.36	0.93
$F_2(50;S=0,G=0)$	1.55	-0.01	0.39	0.39	0.94	1.55	-0.02	0.38	0.39	0.95	1.56	-0.03	0.40	0.40	0.92
$F_2(50;S=1,G=0)$	1.52	-0.01	0.39	0.38	0.93	1.53	-0.02	0.37	0.38	0.94	1.53	-0.03	0.39	0.39	0.92
$F_2(50;S=0,G=1)$	4.42	-0.05	0.66	0.67	0.94	4.44	-0.02	0.72	0.69	0.94	4.49	0.00	0.74	0.72	0.94
$F_2(50;S=1,G=1)$	3.99	-0.06	0.59	0.60	0.95	4.02	-0.03	0.64	0.61	0.94	4.12	-0.02	0.65	0.65	0.95
$F_2(60;S=0,G=0)$	3.14	-0.01	0.73	0.75	0.94	3.14	-0.04	0.72	0.75	0.94	3.14	-0.05	0.75	0.76	0.93
$F_2(60;S=1,G=0)$	3.07	-0.01	0.71	0.73	0.94	3.07	-0.04	0.70	0.73	0.94	3.07	-0.05	0.73	0.74	0.93
$F_2(60;S=0,G=1)$	7.81	-0.05	1.09	1.13	0.94	7.87	-0.06	1.16	1.15	0.94	8.13	0.01	1.23	1.23	0.94
$F_2(60;S=1,G=1)$	6.90	-0.09	0.95	1.00	0.94	6.99	-0.10	1.03	1.01	0.94	7.38	-0.06	1.07	1.09	0.94
$F_2(70;S=0,G=0)$	5.39	0.01	1.23	1.28	0.95	5.39	-0.06	1.23	1.28	0.94	5.41	-0.07	1.29	1.31	0.93
$F_2(70;S=1,G=0)$	5.26	0.00	1.20	1.25	0.95	5.26	-0.07	1.20	1.25	0.94	5.28	-0.08	1.25	1.28	0.93
$F_2(70;S=0,G=1)$	11.38	-0.03	1.66	1.71	0.94	11.57	-0.12	1.69	1.74	0.95	12.34	0.01	1.87	1.93	0.95
$F_2(70;S=1,G=1)$	9.97	-0.10	1.45	1.53	0.95	10.22	-0.19	1.51	1.55	0.95	11.20	-0.10	1.66	1.74	0.94

Table A.6: (250 families) Empirical properties of the parameter and penetrance estimators for shared frailty competing risks model with Cox and Oakes TVC under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1 based on 500 simulations.

	Cox and Oakes Model														
	True	$k_1 = 7$, $\tau = 0.07$				True	$k_1 = 3.5$, $\tau = 0.13$				True	$k_1 = 1$, $\tau = 0.33$			
	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP	value	Bias	ESE	ASE	ECP
$log(\lambda_1)$	-4.83	0.00	0.07	0.08	0.95	-4.83	0.00	0.07	0.08	0.96	-4.83	0.00	0.07	0.09	0.95
$log(\rho_1)$	0.83	0.01	0.04	0.05	0.96	0.83	0.00	0.04	0.05	0.94	0.83	0.01	0.04	0.05	0.95
$log(\lambda_2)$	-4.96	-0.01	0.11	0.13	0.95	-4.96	-0.01	0.11	0.13	0.94	-4.96	0.01	0.10	0.12	0.94
$log(\rho_2)$	1.07	0.00	0.08	0.09	0.94	1.07	0.00	0.07	0.09	0.96	1.07	0.01	0.07	0.08	0.96
β_s	1.52	0.11	0.47	0.61	0.94	1.52	0.05	0.48	0.57	0.93	1.52	0.10	0.48	0.57	0.93
β_{g1}	2.08	0.03	0.14	0.17	0.95	2.08	0.01	0.15	0.17	0.95	2.08	0.01	0.14	0.16	0.96
β_{g2}	1.57	0.03	0.24	0.29	0.96	1.57	0.02	0.24	0.29	0.97	1.57	0.02	0.26	0.30	0.96
$log(k_1)$	1.95	0.26	0.99	1.13	0.89	1.25	0.12	0.58	0.72	0.95	0.00	0.03	0.27	0.32	0.97
$log(k_2)$	1.26	0.57	1.52	1.55	0.82	1.26	0.57	1.44	1.63	0.81	1.26	0.67	1.51	1.76	0.81
$log(\eta)$	-0.18	-0.01	0.71	0.80	0.87	-0.18	-0.07	0.69	0.82	0.87	-0.18	-0.13	0.68	0.77	0.87
η_0	0.21	-0.03	0.17	0.21	0.95	0.21	-0.03	0.17	0.21	0.96	0.21	-0.05	0.19	0.22	0.94
Cause-specific penetrance (%) given $t_{sc} = 35$															
Cause 1															
$F_1(40;S=0,G=0)$	2.43	-0.04	0.39	0.40	0.94	2.42	-0.02	0.41	0.42	0.94	2.40	0.01	0.41	0.43	0.95
$F_1(40;S=1,G=0)$	3.25	0.00	0.58	0.63	0.95	3.24	0.02	0.62	0.65	0.95	3.21	0.10	0.62	0.66	0.95
$F_1(40;S=0,G=1)$	17.45	-0.03	1.73	1.69	0.94	17.23	-0.12	1.70	1.78	0.95	16.27	0.09	1.94	2.01	0.95
$F_1(40;S=1,G=1)$	22.47	0.30	2.82	2.85	0.97	22.12	0.16	2.80	2.93	0.95	20.56	0.54	2.78	3.01	0.98
$F_1(50;S=0,G=0)$	5.15	-0.07	0.77	0.80	0.95	5.13	-0.04	0.79	0.82	0.95	5.04	0.03	0.81	0.85	0.96
$F_1(50;S=1,G=0)$	6.57	-0.02	1.09	1.16	0.94	6.54	0.02	1.15	1.18	0.95	6.39	0.10	1.14	1.20	0.95
$F_1(50;S=0,G=1)$	32.90	0.05	2.77	2.74	0.94	32.16	-0.14	2.72	2.91	0.97	29.06	0.21	3.13	3.29	0.96
$F_1(50;S=1,G=1)$	39.48	0.33	4.15	4.15	0.94	38.43	0.12	4.15	4.26	0.95	34.14	0.42	4.05	4.34	0.95
$F_1(60;S=0,G=0)$	8.91	-0.09	1.27	1.33	0.96	8.85	-0.06	1.31	1.36	0.95	8.58	0.07	1.33	1.39	0.96
$F_1(60;S=1,G=0)$	11.09	-0.12	1.82	1.91	0.93	11.00	-0.08	1.87	1.96	0.95	10.59	-0.01	1.85	1.93	0.96
$F_1(60;S=0,G=1)$	48.48	0.11	3.55	3.65	0.95	46.94	-0.14	3.57	3.88	0.97	40.89	0.30	4.07	4.32	0.96
$F_1(60;S=1,G=1)$	55.53	-0.05	5.06	5.10	0.95	53.57	-0.25	5.05	5.31	0.96	46.03	0.02	4.99	5.31	0.96
$F_1(70;S=0,G=0)$	13.54	-0.09	1.88	2.00	0.96	13.41	-0.07	1.93	2.02	0.95	12.81	0.12	1.94	2.03	0.96
$F_1(70;S=1,G=0)$	16.60	-0.27	2.70	2.85	0.93	16.41	-0.22	2.72	2.89	0.96	15.52	-0.15	2.68	2.80	0.95
$F_1(70;S=0,G=1)$	61.12	0.10	3.87	4.12	0.96	58.82	-0.13	4.06	4.42	0.97	50.11	0.31	4.62	4.91	0.96
$F_1(70;S=1,G=1)$	67.55	-0.40	5.15	5.28	0.95	64.90	-0.50	5.22	5.59	0.96	54.88	-0.26	5.42	5.75	0.96
Cause 2															
$F_2(40;S=0,G=0)$	0.62	0.00	0.18	0.18	0.93	0.62	-0.01	0.17	0.17	0.94	0.62	0.01	0.18	0.18	0.93
$F_2(40;S=1,G=0)$	0.62	0.00	0.18	0.18	0.93	0.62	-0.01	0.17	0.17	0.94	0.62	0.01	0.18	0.18	0.93
$F_2(40;S=0,G=1)$	2.66	0.00	0.45	0.46	0.94	2.67	-0.03	0.45	0.46	0.95	2.68	0.02	0.48	0.48	0.92
$F_2(40;S=1,G=1)$	2.58	-0.01	0.43	0.44	0.94	2.59	-0.03	0.43	0.45	0.96	2.61	0.02	0.46	0.46	0.92
$F_2(50;S=0,G=0)$	1.61	-0.01	0.39	0.40	0.93	1.61	-0.02	0.39	0.40	0.95	1.61	0.03	0.41	0.41	0.93
$F_2(50;S=1,G=0)$	1.59	-0.01	0.39	0.39	0.93	1.59	-0.02	0.38	0.39	0.94	1.59	0.02	0.40	0.40	0.93
$F_2(50;S=0,G=1)$	6.11	-0.02	0.81	0.83	0.94	6.13	-0.06	0.83	0.84	0.95	6.24	0.06	0.89	0.89	0.94
$F_2(50;S=1,G=1)$	5.66	-0.04	0.74	0.77	0.95	5.70	-0.08	0.76	0.78	0.95	5.86	0.03	0.81	0.82	0.94
$F_2(60;S=0,G=0)$	3.23	-0.02	0.73	0.76	0.94	3.23	-0.03	0.74	0.76	0.95	3.24	0.06	0.77	0.79	0.94
$F_2(60;S=1,G=0)$	3.18	-0.02	0.72	0.74	0.94	3.18	-0.04	0.73	0.74	0.95	3.18	0.06	0.76	0.77	0.94
$F_2(60;S=0,G=1)$	10.30	-0.03	1.31	1.34	0.95	10.42	-0.06	1.34	1.36	0.95	10.90	0.13	1.42	1.46	0.95
$F_2(60;S=1,G=1)$	9.22	-0.06	1.19	1.23	0.96	9.37	-0.09	1.22	1.25	0.95	10.00	0.09	1.28	1.34	0.96
$F_2(70;S=0,G=0)$	5.53	-0.03	1.23	1.29	0.94	5.53	-0.04	1.28	1.30	0.94	5.55	0.13	1.31	1.35	0.95
$F_2(70;S=1,G=0)$	5.39	-0.03	1.19	1.26	0.94	5.39	-0.04	1.24	1.27	0.94	5.42	0.13	1.28	1.31	0.94
$F_2(70;S=0,G=1)$	14.27	-0.04	1.89	1.93	0.95	14.61	-0.05	1.96	1.98	0.95	15.91	0.20	2.10	2.19	0.96
$F_2(70;S=1,G=1)$	12.35	-0.03	1.77	1.82	0.95	12.77	-0.04	1.82	1.87	0.94	14.36	0.19	1.95	2.06	0.95

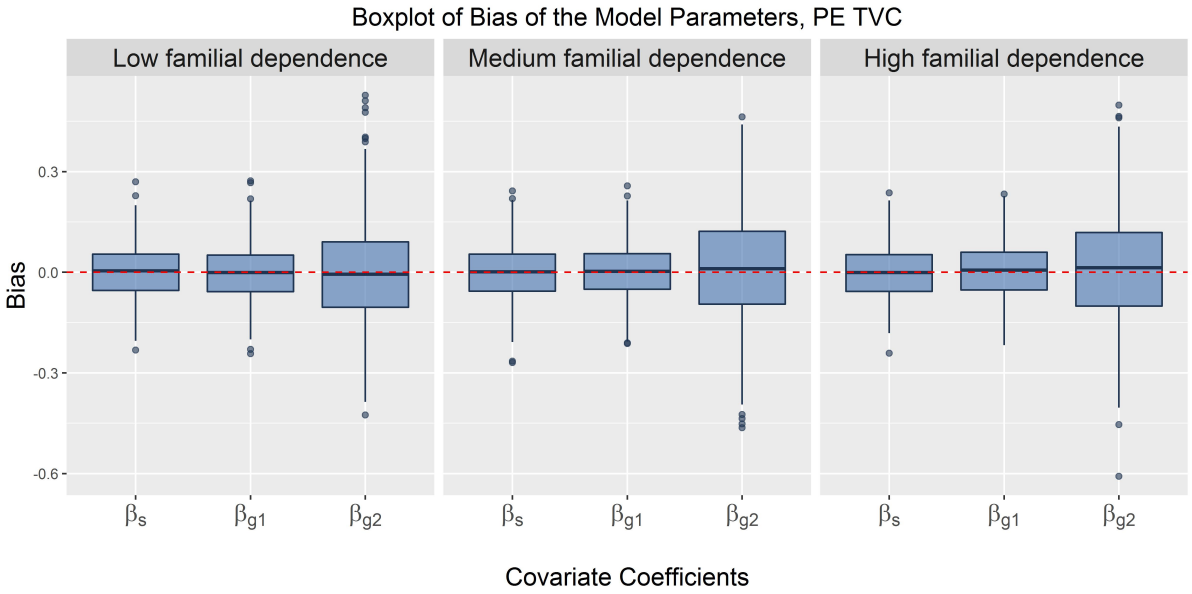

Figure A.1: (1000 families) Boxplot of bias of the model parameters for PE TVC model under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1.

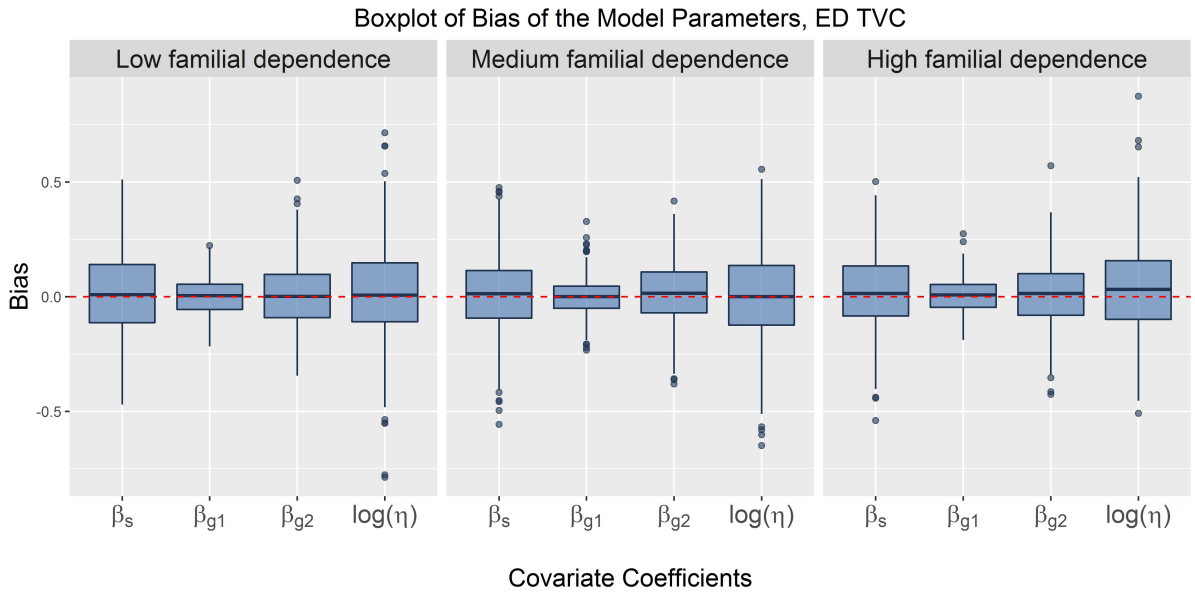

Figure A.2: (1000 families) Boxplot of bias of the model parameters for ED TVC model under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1.

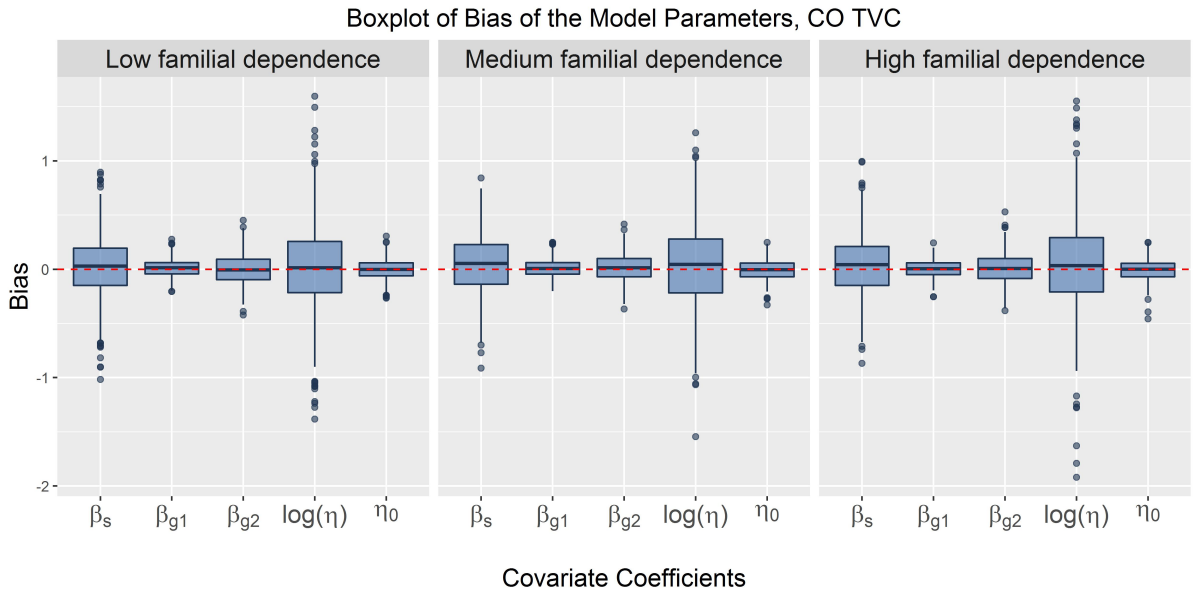

Figure A.3: (1000 families) Boxplot of bias of the model parameters for CO TVC model under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1.

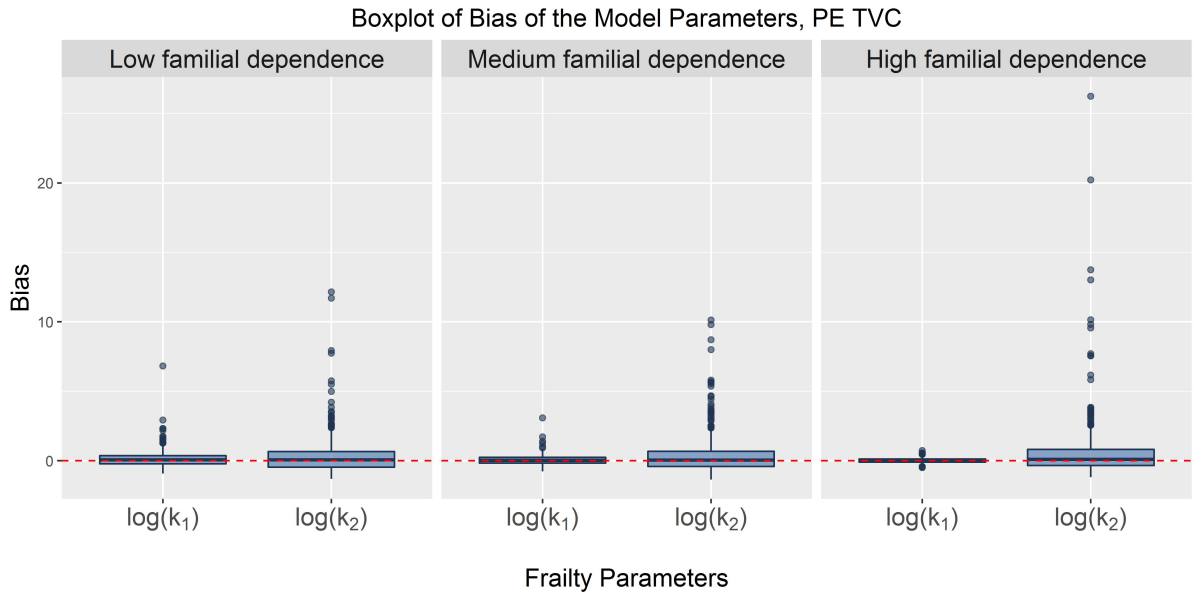

Figure A.4: (1000 families) Boxplot of bias of the frailty parameters for PE TVC model under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1.

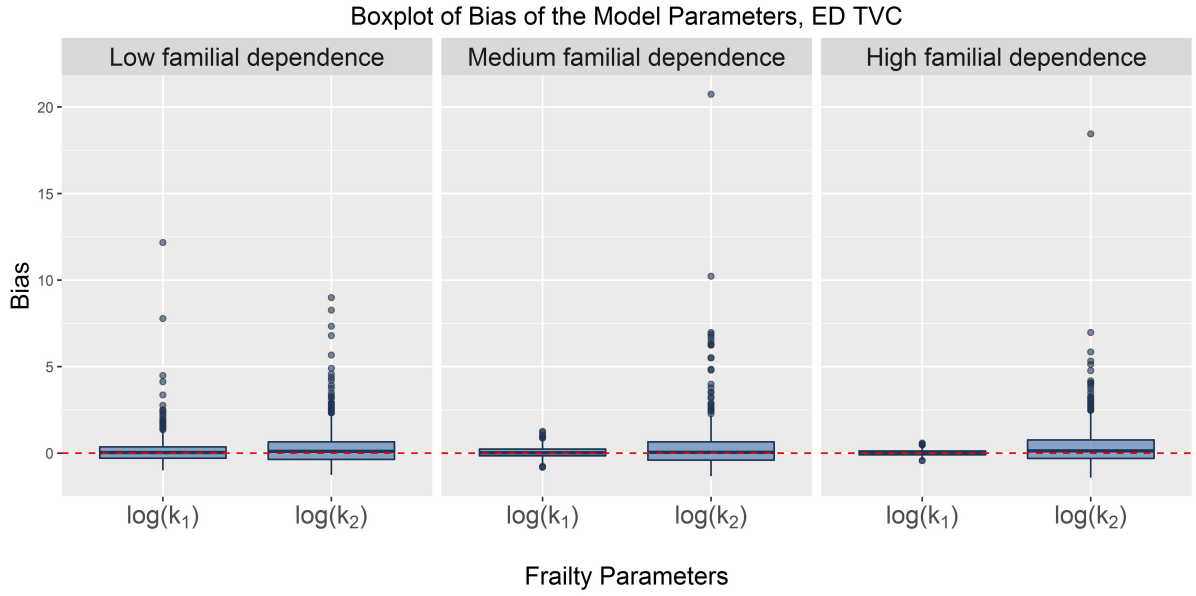

Figure A.5: (1000 families) Boxplot of bias of the frailty parameters for ED TVC model under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1.

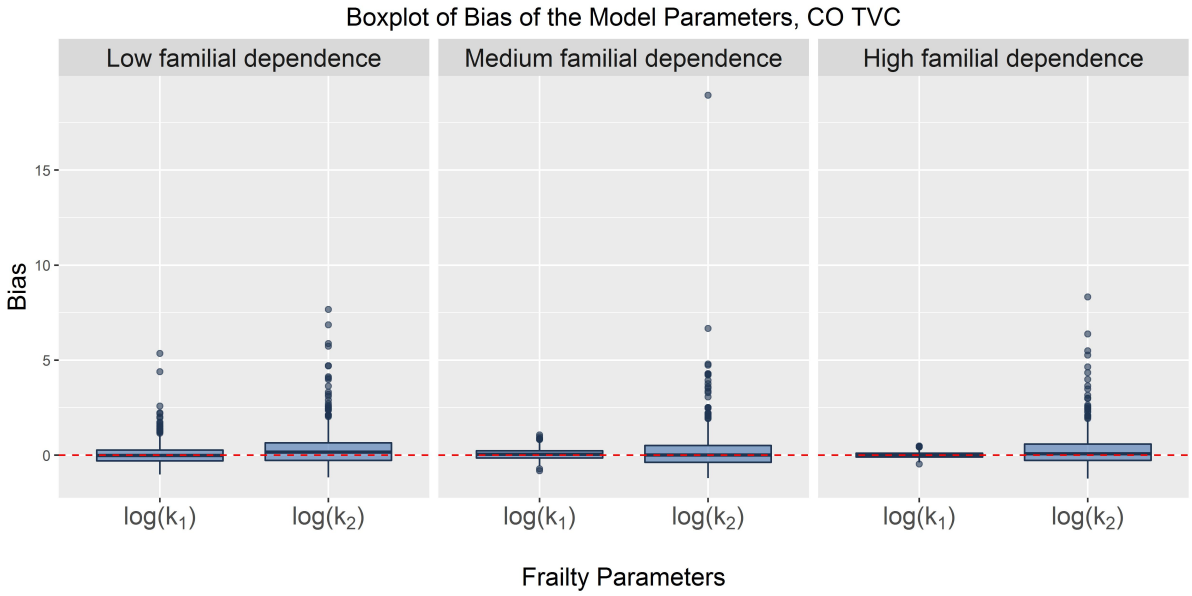

Figure A.6: (1000 families) Boxplot of bias of the frailty parameters for CO TVC model under low ($k_1 = 7$), medium ($k_1 = 3.5$) and high ($k_1 = 1$) familial dependence of the event times for event 1.

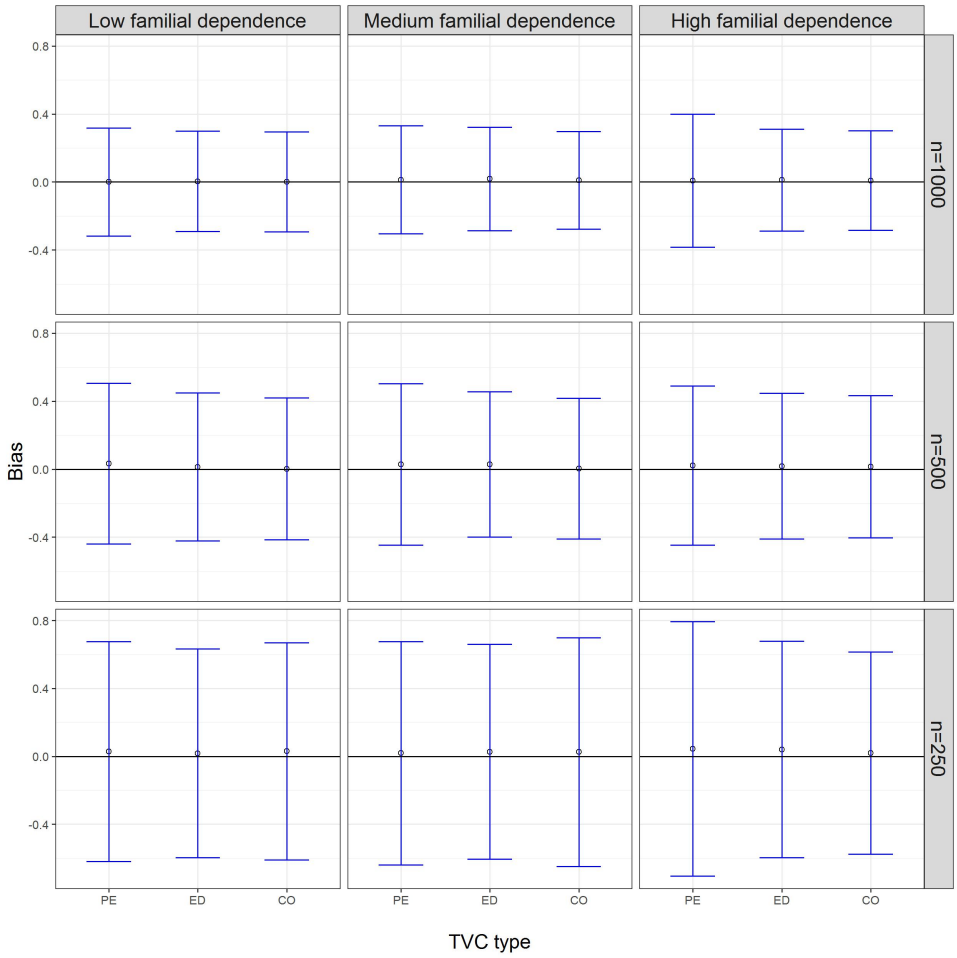

Figure A.7: Bias and precision of the parameter estimates for mutation effect parameter for event 2, $\beta_{g,1}$, expressed as mean +/- 1.96ASE, based on 500 simulations. True value of the $\beta_{g,2}$ is 1.194 for PE, 1.224 for ED and 1.566 for CO.

Appendix B

R codes

```
# FITTING MODEL CODES
FitModelEM <- function( data = data,
                        init.Parms = list(
                          cause1 = list(base = c(0.008,2.581), time_indep=c(2.084), time_dep=list(main=c(2.364,1.033,1.128,0.411)),
                          cause2 = list(base = c(0.006,2.538), time_indep=c(1.437), time_dep=list(main=0.818)),
                          cause3 = list(base = c(0.016,4.193), time_indep=c(-0.094), time_dep=list(main=-1.851))
                         ),

                        time.dep.cov = list(time.dep.cov.type = "PE" , #  "ED")
                                            reccur.type = "IS"  #  "CM")
                        ),

                        missing.method = "data",
                        base.dist = "Weibull",
                        frailty=FALSE,
                        weight=FALSE,
                        mutation.prediction=FALSE
){
  #initial parameters
  base.parms <- c(init.Parms[[1]]$base, init.Parms[[2]]$base, init.Parms[[3]]$base)
  # total 6, lambda1,rho1 lambda2,rho2, lambda3,rho3
  vbeta_b <- c( init.Parms[[1]]$time_indep, init.Parms[[1]]$time_dep$main, init.Parms[[1]]$time_dep$interaction )
  # breast cancer parameter
  vbeta_o <- c( init.Parms[[2]]$time_indep, init.Parms[[2]]$time_dep$main )
  # ovarian cancer parameter
  vbeta_d <- c( init.Parms[[3]]$time_indep, init.Parms[[3]]$time_dep$main )
  # death parameter
  vbeta <- c(vbeta_b,vbeta_o,vbeta_d)
  tvctype=time.dep.cov$time.dep.cov.type
  if(tvctype=="ED"){
    phi_sc=c(log(c(0.5,0.5,0.5)))
    phi_or=c(log(0.5))
  } else if (tvctype=="CO"){
    phi_sc=c(log(c(0.5,0.5,0.5)))
    phi_or=c(log(0.5),0)
```

```
  } else if (tvctype=="PE"){
    phi_sc=c(log(c(0.5,0.5,0.5)))
    phi_or=NULL
  } else if (tvctype=="TIC"){
    phi_sc=c(log(c(0.5,0.5,0.5)))
    phi_or=NULL
  }
  if(frailty==TRUE){
      datacopy = data
      fsize <- aggregate(datacopy$status, by=list(datacopy$famID), length)[,2]
      df1 <- aggregate(datacopy$status==1, by=list(datacopy$famID), FUN=sum)[,2]
      df2 <- aggregate(datacopy$status==3, by=list(datacopy$famID), FUN=sum)[,2]
      #df3 <- aggregate(datacopy$status==4, by=list(datacopy$famID), FUN=sum)[,2]
      data$df1 <- rep(df1, fsize)
      data$df2 <- rep(df2, fsize)
      #data$df3 <- rep(df3, fsize)
      fp <- c(3.5, 3.5)#, 3.5)
  } else {
      fp = NULL
  }
    theta = c(log(base.parms),vbeta,phi_sc,phi_or,log(fp)) # length(theta) = 14
    # data preparation
    data2 <- Data_preparation(data)
    # Imputing Missing mutation status
    data.cooked = carrierprobgeno(method = missing.method, data=data)
    impute=function(dat){
      for(i in 1:nrow(dat)){
        if(is.na(dat$mgene[i]))dat[i,"mgene"]<-sample(c(0,1),1,prob=c(1-dat$carrp.geno[i],dat$carrp.geno[i]),replace=TRUE)
      }
      return(dat)
    }
    data.cooked = impute(data.cooked)
    est0 <- est <- theta
    est1 <- optim(est, loglik_Comp_Timedep, data=data.cooked, data2=data2,
                  agemin=16, frailty=TRUE,
                  tvctype = tvctype,
                  hessian=TRUE, control=list(maxit=10000))
    print(est1$convergence)
    return(est1)
}

# Logliklihood function
loglik_Comp_Timedep=function(theta, data, data2, agemin, frailty=FALSE, tvctype){
  data = data[data$currentage>=agemin,]
  # base parameters
  lambda1 = exp(theta[1]);rho1 = exp(theta[2]);lambda2 = exp(theta[3]);rho2 = exp(theta[4])
  lambda3 = exp(theta[5]);rho3 = exp(theta[6])
  # vbeta for breast
  beta.gen1 = theta[7];beta.sc1_1 = theta[8];beta.sc2_1 = theta[9];beta.sc3_1 = theta[10]
  beta.or1 = theta[11]
  # vbeta for ovarian
  beta.gen2 = theta[12]
  # vbeta for death
  beta.gen3 = theta[13]
```

```
# time varying covariate type
if(tvctype=="ED"){
  phi_sc = c(exp(theta[14:16]),0,0,0) #ED
  phi_or = c(exp(theta[17]),0)
}else if(tvctype=="CO"){
  phi_sc = c(exp(theta[14:16]),0,0,0) #ED
  phi_or = c(exp(theta[17]),theta[18]) #CO
}else if(tvctype=="PE"){
  phi_sc = c(exp(theta[14:16]),0,0,0) #ED
  phi_or = c(0,0)
}else if(tvctype=="TIC"){
  phi_sc = c(exp(theta[14:16]),0,0,0)
  phi_or = c(0,0)
}
# frailty parameter
if(frailty==TRUE){
  fp <- exp(theta[(length(theta)-1):length(theta)])
}else{
  fp=NULL
}
# Y, delta, carrier prob, proband indicator, sc number
time0 = data$time-agemin
status = data$status
ip = which(data$proband==1)
mgene=data$mgene
sc_num = data$screen_number
sc_num_p = data$screen_number_p
#setting up indicator variable for the removal status
br_time <- data$br.censortime-agemin
delta_BR <- ifelse(time0-br_time > 0,1,0)
sc1_time <- data$st1 - agemin
sc2_time <- data$st2 - agemin
sc3_time <- data$st3 - agemin
Z_SC1=  ifelse(is.na(sc1_time), 0,1)
Z_SC2=  ifelse(is.na(sc2_time), 0,1)
Z_SC3=  ifelse(is.na(sc3_time), 0,1)
# Variables related to oophorectomy
 ov_time <- data$ov.censortime-agemin
delta_OR <- ifelse(time0-ov_time > 0,1,0)
    Z_OR <- rep(1,nrow(data))
if(tvctype!="TIC"){
   Z_OR = ifelse( delta_OR==0, 0, ZED(time0, ov_time, phi_or[1]))
  Z_SC1 = ifelse( Z_SC1==0, 0, ZED(time0, sc1_time, phi_sc[1]))
  Z_SC2 = ifelse( Z_SC2==0, 0, ZED(time0, sc2_time, phi_sc[2]))
  Z_SC3 = ifelse( Z_SC3==0, 0, ZED(time0, sc3_time, phi_sc[3]))
}
#loading up event time(sorted) vectors, event indicator vectors
time_vec2 <- data2[[1]]
indicator_vec <- data2[[2]]
time_vec2_p0 <- data2[[3]]
indicator_vec_p0 <- data2[[4]]
time_vec2_p1 <- data2[[5]]
indicator_vec_p1 <- data2[[6]]
time_vec2_p2 <- data2[[7]]
```

```
 indicator_vec_p2 <- data2[[8]]
 # baseline hazard(weibull), baseline cumulative hazard(weibull)
 bhaz1 = (lambda1^rho1)*rho1*time0^(rho1-1)
 bcumhaz1 = (lambda1*time0)^rho1
 bhaz2 = (lambda2^rho2)*rho2*time0^(rho2-1)
 bcumhaz2 = (lambda2*time0)^rho2
 bhaz3 = (lambda3^rho3)*rho3*time0^(rho3-1)
 bcumhaz3 = (lambda3*time0)^rho3
 logh1 = log(bhaz1) + mgene*beta.gen1 ###
 logh2 = log(bhaz2) + mgene*beta.gen2
 logh3 = log(bhaz3) + mgene*beta.gen3
 # creating screen,gene,oophorectomy interaction vectors for hazard
 sc_vec <-rep(0,nrow(data))
 sc_vec[sc_num==0] <- 0
 sc_vec[sc_num!=0] <- sc_num[sc_num!=0]
 sc_vec_bc  <- vector("numeric", length(sc_num) )
 sc_vec_bc[sc_vec==0] <- 0
 sc_vec_bc[sc_vec==1] <- Z_SC1[sc_vec==1]*beta.sc1_1 + phi_sc[4]
 sc_vec_bc[sc_vec==2] <- Z_SC2[sc_vec==2]*beta.sc2_1 + phi_sc[5]
 sc_vec_bc[sc_vec==3] <- Z_SC3[sc_vec==3]*beta.sc3_1 + phi_sc[6]
 # sum of log-hazard
 sum1 = sum((logh1+ (delta_OR*Z_OR*beta.or1 + delta_OR*phi_or[2]) + sc_vec_bc)[status==1], na.rm=TRUE)
 if(tvctype=="TIC"){
 sum1 = sum((logh1+ (delta_OR*beta.or1) + sc_vec_bc)[status==1], na.rm=TRUE)
 }
 sum2 = sum(logh2[status==3], na.rm=TRUE)
 sum3 = sum(logh3[status==4], na.rm=TRUE)
 # sum of survival until event time for all the individuals in the data
 cest <- c(lambda1,rho1,lambda2,rho2,lambda3,rho3,beta.gen1,beta.sc1_1,beta.sc2_1,
           beta.sc3_1,beta.or1,beta.gen2,beta.gen3)
mat_all <- cbind(time_vec2, indicator_vec, time0, mgene)
        if(tvctype \%in\% c("PE","ED","CO")){
         testm <- mat_p1
         testm[is.na(testm)] <- 99
         CH_bc=CumH_c_CO(v1= testm[,1:5], v2=testm[,6:10], u = testm[,11],cest=cest, affp=FALSE, type=1,phi=phi_sc,phi2=phi_or)
         CH_ov=CumH_c_CO(v1= testm[,1:5], v2=testm[,6:10], u = testm[,11],cest=cest, affp=FALSE, type=2,phi=phi_sc,phi2=phi_or)
         CH_d=CumH_c_CO(v1= testm[,1:5], v2=testm[,6:10], u = testm[,11],cest=cest,  affp=FALSE, type=3,phi=phi_sc,phi2=phi_or)
       }
        if(tvctype=="TIC"){
        CH_bc <-    exp(testm[,12]*beta.gen1+delta_OR*beta.or1)*CH_bc[,1]
        }else{
        CH_bc <-    exp(testm[,12]*beta.gen1)*CH_bc[,1]
        }
        CH_ov <-    exp(testm[,12]*beta.gen2)*CH_ov[,1]
        CH_d  <-    exp(testm[,12]*beta.gen3)*CH_d[,1]
          k <- CH_bc + CH_ov + CH_d
        sum4=k
 if(frailty==TRUE){ # Gamma-frailty
   Hfam1 <- -CH_bc
   Hfam2 <- -CH_ov
   Hfam3 <- -CH_d
   df1 <- data$df1[data$proband==1]
   df2 <- data$df2[data$proband==1]
   df3 <- data$df3[data$proband==1]
```

```
  Hfam1 <- aggregate(Hfam1,by=list(data$famID),FUN=sum)[,2]
  Hfam2 <- aggregate(Hfam2,by=list(data$famID),FUN=sum)[,2]
  Hfam3 <- aggregate(Hfam3,by=list(data$famID),FUN=sum)[,2]
  sum4.1 <- sum((lfactorial(fp[1]+df1-1)-(df1-1)*log(fp[1])-lfactorial(fp[1]) +
                (-fp[1]-df1)*log(1+(Hfam1)/fp[1])), na.rm=T)
  sum4.2 <- sum((lfactorial(fp[2]+df2-1)-(df2-1)*log(fp[2])-lfactorial(fp[2]) +
                (-fp[2]-df2)*log(1+(Hfam2)/fp[2])), na.rm=T)
  #sum4.3 <- sum((lfactorial(fp[3]+df3-1)-(df3-1)*log(fp[3])-lfactorial(fp[3]) +
                (-fp[3]-df3)*log(1+(Hfam3)/fp[3])), na.rm=T)
  loglik = sum1+sum2+sum3+sum4.1+sum4.2+sum(Hfam3,na.rm=TRUE)
}else{
  sum4 <- sum(sum4,na.rm = TRUE)
  loglik = sum1+sum2+sum3+sum4 # numerator in loglikelihood
}
# Ascertainment correction by design="pop+"
cagep <- data$currentage[ip]-agemin
timep <- data$time[ip]-agemin
statusp <- data$affect[ip]  # proband disease status at the study entry
# subsetting the probands according to his affection status at study entry
cagep0 <- cagep[statusp==0]
timep0 <- timep[statusp==0]
mat_p0 <- cbind(time_vec2_p0, indicator_vec_p0, cagep0)
    if(tvctype \%in\% c("PE","ED","CO")){
      testm <- mat_p1
      testm[is.na(testm)] <- 99
      CH_bc=CumH_c_CO(v1= testm[,1:5], v2=testm[,6:10], u = testm[,11],cest=cest, affp=FALSE, type=1,phi=phi_sc,phi2=phi_or)
      CH_ov=CumH_c_CO(v1= testm[,1:5], v2=testm[,6:10], u = testm[,11],cest=cest, affp=FALSE, type=2,phi=phi_sc,phi2=phi_or)
      CH_d=CumH_c_CO(v1= testm[,1:5], v2=testm[,6:10], u = testm[,11],cest=cest,  affp=FALSE, type=3,phi=phi_sc,phi2=phi_or)
    }
    if(tvctype=="TIC"){
      CH1_bc <- exp(beta.gen1+delta_OR[ip][statusp==0]*beta.or1)*CH_bc[,1]
    }else{
      CH1_bc <- exp(beta.gen1)*CH_bc[,1]
    }
    CH1_ov <- exp(beta.gen2)*CH_ov[,1]
    CH1_d <-  exp(beta.gen3)*CH_d[,1]
    logasc0 <- CH1_bc + CH1_ov +CH1_d
if(frailty==TRUE){
  logS <- log((1-CH1_bc/fp[1])^(-fp[1])) + log((1-CH1_ov/fp[2])^(-fp[2])) +  CH1_d
}else {
  logS <- logasc0
}
logasc0 <- sum(logS,na.rm=TRUE)
# subsetting the probands according to his affection status at study entry
cagep1 <- cagep[statusp==1|statusp==2]
timep1 <- timep[statusp==1|statusp==2]
mat_p1 <- cbind(time_vec2_p1, indicator_vec_p1, cagep1)
    if(tvctype \%in\% c("PE","ED","CO")){
      testm <- mat_p1
      testm[is.na(testm)] <- 99
      CH_bc=CumH_c_CO(v1= testm[,1:5], v2=testm[,6:10], u = testm[,11],cest=cest, affp=FALSE, type=1,phi=phi_sc,phi2=phi_or)
      CH_ov=CumH_c_CO(v1= testm[,1:5], v2=testm[,6:10], u = testm[,11],cest=cest, affp=FALSE, type=2,phi=phi_sc,phi2=phi_or)
      CH_d=CumH_c_CO(v1= testm[,1:5], v2=testm[,6:10], u = testm[,11],cest=cest,  affp=FALSE, type=3,phi=phi_sc,phi2=phi_or)
    }
```

```
      if(tvctype=="TIC"){
        CH1_bc <- exp(beta.gen1+delta_OR[ip][statusp==1|statusp==2]*beta.or1)*CH_bc[,1]
      }else{
        CH1_bc <- exp(beta.gen1)*CH_bc[,1]
      }
        CH1_ov <- exp(beta.gen2)*CH_ov[,1]
        CH1_d <-  exp(beta.gen3)*CH_d[,1]
  if(frailty==TRUE){
    logS <- log(1-((1-CH1_bc/fp[1])^(-fp[1]))*((1-CH1_ov/fp[2])^(-fp[2]))*exp(CH1_d))
  }else {
    logS <- log(1-exp(CH1_bc+CH1_ov) )
  }
  logasc12=sum(logS,na.rm=TRUE)
  slogasc = logasc0 + logasc12
  likelihood  <- loglik - slogasc
  return(-likelihood)
}

#C++ and R integrated function via Rcpp package for numerical integration of hazard function
#from the matrix of individuals timeline information of Screens and surgeries history
NumericMatrix CumH_c_CO(NumericMatrix v1, NumericMatrix v2, NumericVector u, NumericVector cest,
                        bool affp, int type, NumericVector phi, NumericVector phi2) {
  int n = v1.nrow();
  NumericMatrix Hvec(n,1);
  NumericVector uv = u;
  for(int j=0; j < n; ++j) {
  double finalH=0;
  double phisc=0, phior=0, phior0=0, phisc0=0;
    double u = uv[j];
    NumericVector k1 = v1( j , _ );
    NumericVector k2 = k1[k1<u];
    int size = k2.size();
    if (size==0) {
      // double chaz(double lam1, double rho1, double betasc, double betaor, double phisc, double phior,
      //             double lowsc, double lowor, double low, double upper)
      if(type==1){
        finalH = -chaz(cest[0],cest[1],0,0,0,0, 0,0, 0,u);
      } else if (type==2){
        finalH = -chaz(cest[2],cest[3],0,0,0,0, 0,0, 0,u);
      } else if (type==3){
        finalH = -chaz(cest[4],cest[5],0,0,0,0, 0,0, 0,u);
      }
    } else {
      k2.push_back(u);
      size = k2.size();
      NumericVector w2 = v2( j , _ );
      NumericVector Hfull;
      double beta_sc=0, beta_or=0, lowsc=0, lowor=0;
      for (int i=0; i < size; ++i) {
        double beta;
        if(i!=0){
          beta = coeff_c_CO(w2[i-1],type=type, affp, cest);
          if(w2[i-1]==1 || w2[i-1]==2 || w2[i-1]==3 ){
            beta_sc = beta;
```

```
            phisc = phi[w2[i-1]-1];
            phisc0 = phi[w2[i-1]+2];
            lowsc = k2[i-1];
          }else if(w2[i-1]==4){
            beta_or = beta;
            lowor = k2[i-1];
            if(type==1){
              phior=phi2[0];
              phior0=phi2[1];
            }
          }
        }else{
          beta = 0;
          beta_sc = beta;
          beta_or = beta;
        }
        if(beta==99){
          break;
        }
        double Hpiece, H_noncarrier, low, upper;
        upper = k2[i];
        if(i!=0){
          low = k2[i-1];
        } else{
          low = 0;
        }
        // double chaz(double lam1, double rho1, double betasc, double betaor, double phisc, double phior,
        //             double lowsc, double lowor, double low, double upper)
        if(type==1){
          Hpiece = chaz(cest[0],cest[1],beta_sc,beta_or,phisc,phior,lowsc=lowsc,lowor=lowor, low=low,upper=upper);
          Hpiece = Hpiece*exp(phisc0+phior0);
        } else if (type==2){
          Hpiece = chaz(cest[2],cest[3],0,0,0,0, 0,0, low=low,upper=upper);
        } else {
          Hpiece = chaz(cest[4],cest[5],0,0,0,0, 0,0, low=low,upper=upper);
        }
        H_noncarrier = Hpiece;
        Hfull.push_back(H_noncarrier);
      }
      finalH = -sum(Hfull);
    }
    Hvec(j,0) = finalH;
  }
  return Hvec;
}
```

Bibliography

[1] Aalen, O.O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous markov chains based on censored observations. *Scandinavian Journal of Statistics*, 5:141-150.

[2] Beyersmann, J., Latouche, A., Buchholz, A., and Schumacher, M. (2009). Simulating competing risks data in survival analysis. *Statistics in Medicine*, 28:956-971.

[3] Burton, A., Altman, D., Royston, P., and Holder, R. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, 25:4279-4292.

[4] Choi, Y.-H. (2012). A frailty-model based method for estimating age dependent penetrance from family data. *Journal of Biometrics and Biostatistics*, S4:001.

[5] Choi, Y.-H., Kopciuk, K., and Briollais, L. (2008). Estimating disease risks associated with mutated genes in family-based designs. *Human Hereditary*, 66:238-251.

[6] Choi, Y.-H., Kopciuk, K., He, W., and Briollais, L. (2017). FamEvent: family age-at-onset data simulation and penetrance estimation. R package version 1.3.

[7] Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65:141-151.

[8] Cox, D.R., Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall, New York.

[9] Daniel, S., Koren, G., Lunenfeld, E., and Levy, A. (2015). Immortal time bias in drug safety cohort studies: spontaneous abortion following nonsteroidal anti-inflammatory drug exposure. *American Journal of Obstetrics and Gynecology*, 212:307.

[10] Duchateau, L. and Janssen, P. (2008). *The Frailty Model*. Springer, New York.

[11] Eisen, A., Lubinski, J., Klijn, J., Moller, P., Lynch, H., and Offit, K. Narod, A. (2005). Breast cancer risk following bilateral oophorectomy in BRCA1 and BRCA2 mutation carriers: an international case-control study. *Journal of Clinical Oncology*, 23:7491-7496.

[12] Fine, J. and Gray, R. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94:496-509.

[13] Geskus, R.B. (2016). *Data Analysis with Competing Risks and Intermediate States*. Chapman and Hall, New York.

[14] Goethals, K., Janssen, P., and Duchateau, L. (2008). Frailty models and copulas: similarities and differences. *Journal of Applied Statistics*, 35:1071-1079.

[15] Gong, G. and Whittemore, A. (2003). Optimal designs for estimating penetrance of rare mutations of a disease-susceptibility gene. *Genetic Epidemiology*, 24:173-180.

[16] Gorfine, M. and Hsu, L. (2011). Frailty-based competing risks model for multivariate survival data. *Biometrics*, 67:415-426.

[17] Hougaard, P. (1984). Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika*, 71:75-83.

[18] Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer, New York.

[19] Ibrahim, J., Chu, H., and Chen, L. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28:2796-2801.

[20] John, E., Hopper, J., Beck, J., Knight, J., Neuhausen, S., Senie, R.... Ziogas, A. (2004). The breast cancer family registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast Cancer Research*, 6:R375-R389.

[21] Kalbfleisch, J.D. and Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley and Sons, New York.

[22] Keiding, N., Andersen, P., and Klein, J. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine*, 16:215-224.

[23] Keown-Stoneman, C., Horrocks, J., and Darlington, G. (2018). Exponential decay for binary time-varying covariates in Cox models. *Statistics in Medicine*, 37:776-788.

[24] Koller, M.T., Raatz, H., Steyerberg, E.W., and Wolbers, M. (2012). Competing risks and the clinical community: irrelevance or ignorance? *Statistics in Medicine*, 31:1089-97.

[25] Laden, F. and Hunter, D.G. (1998). Environmental risk factors and female breast cancer. *Annual Review of Public Health*, 19:101-123.

[26] Lai, X., Yau, K.W., and Liu, L. (2017). Competing risk model with bivariate random effects for clustered survival data. *Computational Statistics and Data Analysis*, 112:215-223.

[27] Larson, M. and Dinse, G. (1985). A mixture model for the regression analysis of competing risks data. *Applied Statistics*, 34:201-211.

[28] Lee, E., Wei, L., David, A., and Amato, S. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. *Survival Analysis: State of the Art*, 12:237-247.

[29] Lee, M., Ha, I.D., and Lee, Y. (2017). Frailty modeling for clustered competing risks data with missing cause of failure. Statistical Methods in Medical Research, 26:356-373.

[30] Lunn, M. and McNeil, D. (1995). Applying Cox regression to competing risks. *Biometrics*, 51:524-532.

[31] Munda, M., Rotolo, F., and Legrand, C. (2012). parfm: Parametric frailty models in R. *Journal of Statistical Software*, 51:1-20.

[32] Ng, S.K. and McLachlan, G. (2003). An EM-based semi-parametric mixture model approach to the regression analysis of competing risks data. *Statistics in Medicine*, 22:1097-1111.

[33] Noordzij, M., Leffondre, K., and van Stralen, K.J. (2013). When do we need competing risks methods for survival analysis in nephrology? *Nephrol Dial Transplant*, 28:2670-2677.

[34] Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society*, 44:414-422.

[35] Petrucelli, N., Daly, M.B., and Pal, T. (2016, December). BRCA1- and BRCA2-associated hereditary breast and ovarian cancer. *GeneReviews*. Retrieved from: `https://www.ncbi.nlm.nih.gov/books/NBK1247/`.

[36] Prentice, R.L, Kalbfleisch, J.D, Peterson, A., Fluornoy, N., Farewell, V., and Breslow, N. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 34:541-554.

[37] Rizopoulos, D. (2012) *Joint Models for Longitudinal and Time-to-Event Data*. Chapman and Hall, New York.

[38] Schafer, J. and Graham, J. (2002). Missing data: our view of the state of the art. *Psychological Methods*, 7:147-177.

[39] Self, S. and Liang, K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82:605-610.

[40] Southern, D.A., Faris, P.D., Brant, R., Galbraith, P.D., Norris, C.M., Knudtson, M.L., and Ghali, W.A. (2006). Kaplan-Meier methods yielded misleading results in competing risk scenarios. *Journal of Clinical Epidemiology*, 59:1110-1114.

[41] Spiekerman, C. and Lin, D. (1998). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association*, 93:1164-1175.

[42] Tai, B., Machin, D., White, I., and Gebski, V. (2001). Competing risks analysis of patients with osteosarcoma: a comparison of four different approaches. *Statistics in Medicine*, 20:661-684.

[43] Tang, L.Q., Song, J.W., Belin, T.R., and Unutzer, J. (2005). A comparison of imputation methods in a longitudinal randomized clinical trial. *Statistics in Medicine*, 24:2111-2128.

[44] Vaupel, J.W., Manton, K.G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16:439-454.

HAE YOUNG (JAY) JUNG

EDUCATION

MS	University of Western Ontario, Biostatistics	2018
BA	University of Guelph, Economics Minored in Statistics	2016

RESEARCH EXPERIENCE

Research Assistant, University of Western Ontario,
Shulich School of Medicine & Dentistry, London, ON — Sep 2016-current
Supervisor: Dr. Yun-hee Choi

Summer Research Student, Mount Sinai Hospital,
Lunenfeld-Tanenbaum Research Institute, Toronto, ON — May 2017 to Aug 2017
Supervisor: Dr. Laurent Briollais

Research Intern, University of Chicago,
Center for Health Statistics, Chicago, IL — May 2015 to Aug 2015
Supervisor: Dr. Robert Gibbons

COMPUTING SKILLS

Statistical Analysis: R, SAS, STATA
Machine Learning: Python
Programming technologies: SQL
Programming technologies, front end: LaTeX, HTML
Presentation: Adobe Photoshop, Autodesk 3ds Max, Microsoft Words, PowerPoints, Excel

PUBLICATIONS

Journal Papers in Review

Jung, H.Y., Choi, Y.H., Tounkara, F., Knight, J., Andrulis, I., and Briollais, L., "A shared frailty competing risk model with time varying covariates for estimation of breast and ovarian cancer risks in BRCA1/2 families" Submitted to: Western Journal of Graduate Research.